Bedingungen schulischer Anstrengungsbereitschaft

Test von Kausalmodellen an Schülern einer Förderstufe

Peter Sprave

aus Berlin

1979

Bedingungen schulischer Anstrengungsbereitschaft

Test von Kausalmodellen an Schülern einer Förderstufe

INAUGURAL-DISSERTATION

zur Erlangung der Doktorwürde
des
Fachbereichs Psychologie
der
PHILIPPS-UNIVERSITÄT
Marburg/Lahn

vorgelegt von
PETER SPRAVE
aus Berlin

Marburg/Lahn 1979

UNIV.
LEIDEN
BIBL.

Vom Fachbereich Psychologie der Philipps-Universität
Marburg/Lahn als
Dissertation angenommen am ..20.7.1979.....
Tag der mündlichen Prüfung am ..28.9.1979.....
Berichterstatter ..Prof. Dr. F. Merz.....
Mitberichterstatter ..Prof. Dr. L. Tent.....

Vorbemerkung

Ich möchte an dieser Stelle Herrn G. Schmidtchen für seine wohlwollende Unterstützung dieser Arbeit sehr herzlich danken und ebenso Herrn F. Merz, dessen kritische Anmerkungen zu dieser Arbeit mir sehr geholfen haben. Auch danke ich meinen Kollegen in Zürich, Marburg und Giessen, die mir in vielen Einzeldiskussionen Anregungen und Hinweise vermittelten, so vor allem Herrn W. Otto und Herrn H. Gutscher.
Dank schulde ich ferner Herrn G. Rutz vom hessischen Kultusministerium, Herrn O. Ringsdorf vom Schulaufsichtsbereich Wetzlar, dem Schulleiter und den zuständigen Elternbeiräten, die mir zur Durchführung dieser Untersuchung "grünes Licht" gaben. Nicht unerwähnt soll bleiben die dankenswerte Mitarbeit von Herrn H. Ophoff, der einige seiner Giessener Kollegen zum Mitmachen als Versuchsleiter überreden konnte. Ihnen allen sei hier recht herzlich gedankt.

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41546011_0057

Inhalt

1. EINFUEHRUNG 1

1.1. Kompensatorische Förderung der Anstrengungsbereitschaft in der Förderstufe 1

1.2. Die Institution "Schule" als Bedingung für die Unterstützung durch den Lehrer und für die Anstrengungsbereitschaft des Schülers 14

1.3. Hypothesen dieser Untersuchung, noch nicht prüfbar und noch unvollständig 23

2. METHODIK 29

2.1. Modellspezifikation 29

2.1.1. Vorbemerkungen 29

2.1.2. Auswahl der Versuchspersonen 31

2.1.3. Kontrollvariable 33
2.1.3.1. Unterstützungsverhalten der Eltern 33
2.1.3.2. Geschlecht 35

2.1.4. Erfassung der Variablen 37

2.1.5. Kausale Ordnung der Variablen 49

2.1.6. Kontrolle des systematischen Messfehlers 55

2.1.7. Kontrolle der Repräsentativität von Messungen 62

2.1.8. Hypothesen dieser Untersuchung, prüfbar und vollständig: LISREL-Modellspezifikation 64

2.2. Analyseverfahren 75

2.2.1. Identifizierbarkeit, Schätzen und Testen 75

2.2.2. Interne Validitäten und Messfehler-Varianzen der Einzelindikatoren 78

2.2.3. Kontrollvariable "Geschlecht" 81

2.2.4. Messniveau und Analyseverfahren 85

2.3. Durchführung 86

2.3.1. Ablauf der Untersuchung 86

2.3.2. Versuchspersonen 90

3. ERGEBNISSE 95

3.1. Voruntersuchungen 95

3.1.1. Index-Bildung 95

3.1.2. Interne Validitäten und Messfehler-Varianzen der Einzelindikatoren 101
3.1.2.1. Intelligenztest (CFT2) 101
3.1.2.2. Anstrengungsvermeidungstest (AVT) 104
3.1.2.3. Index der sozialen Herkunft (BRSS) 110

3.1.3. Linearität und Additivität der Indikatorenbeziehungen 113

3.1.4. Test der Messfehler-Hypothesen 115

3.2. Test der Alternativen zur Vermittlungshypothese und Test der Zielsetzung der Förderstufe 120

3.3. Ergänzende Bemerkungen 125

4. ZUSAMMENFASSUNG UND DISKUSSION DER ERGEBNISSE 127

4.1. Ergebnisse 127

4.2. Zur Gültigkeit und Bedeutung der Ergebnisse ... 128

4.3. Pädagogische Folgerungen 134

5. ZUSAMMENFASSUNG 141

6. LITERATURVERZEICHNIS 142

ANHANG 150

Anhang 1: Instruktionen und Items 150
Anhang 2: Informationen zu Daten 169

1. EINFUEHRUNG

Diese Untersuchung beschäftigt sich mit der Frage, ob die an der Förderstufe in Hessen praktizierte äussere Leistungsdifferenzierung zum Abbau herkunftsspezifischer Ungleichheiten in der Anstrengungsbereitschaft (Lern- und Leistungsbereitschaft) der Schüler beiträgt.

1.1. Kompensatorische Förderung der Anstrengungsbereitschaft in der Förderstufe

Spätestens seit Anfang dieses Jahrhunderts ist der Gedanke der optimalen Förderung jedes einzelnen Kindes immer stärker in den Mittelpunkt pädagogischer Verantwortung der Schule gerückt worden und seit den Sechziger Jahren hat er durchwegs Eingang in die bildungspolitischen Programme der politischen Parteien und in die Schulgesetze gefunden. Im Förderungsgedanken ist die Forderung enthalten, dass die Schule für sozial gerechte Zuteilung von Berufs- und Lebenschancen zu sorgen habe. Unter diesem Gesichtspunkt ist es die wichtigste Aufgabe der Schule, allen Schülern zu einem gewissen Ausmass an Anstrengungsbereitschaft zu verhelfen. Denn es ist vor allem diese Anstrengungsbereitschaft, die darüber entscheidet, ob das Kind die von der Schule angebotenen Lernchancen nutzt oder nicht nutzt. Die Anstrengungsbereitschaft gilt als eine wichtige Vorbedingung des Schulerfolgs und zugleich als eine vom Lehrer mehr oder weniger gut beeinflussbare Bedingung.

In den vergangenen Jahren wurden einige Schulversuche gestartet mit dem Ziel, auf die zur Zeit des Uebergangs auf weiterführende Schulen vorfindbaren Unterschiede in den Lernvoraussetzungen schulorganisatorisch angemessener zu reagieren. Es galt, schullaufbahnrelevante Entscheidungen möglichst aufzu-

schieben, um mehr Spielraum für die Förderung zu gewinnen. Die auf die vier Grundschuljahre aufbauenden Förderstufen in Hessen sind Beispiele für diese Bemühungen.

Für die Förderstufe ist die Kombination von integriertem mit differenziertem Unterricht charakteristisch. In den Förderstufen werden die Schüler so zu differenzieren versucht, dass möglichst gute Lernerfolge zustande kommen. Am verbreitetsten ist die Differenzierung nach Leistung in den traditionellen Hauptfächern Mathematik und Sprachen (meist Englisch, mitunter auch Deutsch und/oder Latein). In diesen Fächern werden die Schüler in relativ homogen zusammengesetzten Leistungs- oder Kursniveaugruppen unterrichtet (differenzierte Form der Unterrichtsorganisation). Andere Fächer werden in Kerngruppen unterrichtet, denen Schüler aller Leistungsniveaus angehören (integrierte Form der Unterrichtsorganisation). Die theoretische Begründung für die Einteilung der Schüler in relativ homogene Leistungsgruppen besteht in der Annahme, dass in einer Gruppe von Schülern mit ähnlicher Leistungsfähigkeit bessere Lernerfolge erzielt werden können als in (leistungs-) heterogen zusammengesetzten Gruppen.

Zu Beginn der Förderstufe erfolgt die Zuweisung der Schüler zu den Leistungsgruppen für jedes Fach gesondert, fachspezifisch. Der gleiche Schüler kann also in verschiedenen Haupt- oder "Leistungs"-Fächern unterschiedlichen Kursniveaus angehören. Die Zuweisungen werden meistens halbjährlich überprüft bzw. korrigiert. Die Erstzuweisungen am Ende der vierten Grundschulklasse zu den verschiedenen Kursniveaus (wir gehen im folgenden vom A-B-C-System aus: A = höchstes Kursniveau, B = mittleres Kursniveau, C = niedrigstes Kursniveau) in Mathematik und Sprachen erfolgen meistens nach Ergebnissen eines Schulleistungs- oder Intelligenz-Tests, nach dem Wunsch der Eltern, nach den bisherigen Schulleistungen bzw. nach dem Lehrerurteil (vgl. z.B. BACH, 1972, S. 145).

Es ist unbestritten, dass unter diesen Kriterien einer Kurszuweisung zu Beginn der Förderstufe eine Gleichheit schulischer Startchancen nicht gegeben ist. Die Testintelligenz, der Elternwunsch, die bisher gezeigten Schulleistungen, die zum Einteilungskriterium der Niveaugruppen-Bildung gemacht werden, sind zu einem gewissen Ausmass von der sozialen Herkunft der Schüler bedingt. Es liegen eine ganze Reihe von Ergebnissen vor, die in diese Richtung weisen: Je höher der soziale Status des Elternhauses (üblicherweise gemessen an den Statusfaktoren Beruf, Bildungsstand und Einkommen der Eltern), desto anspruchsvoller ist der Elternwunsch in Bezug auf die Schulausbildung, desto besser sind im allgemeinen auch der Schulerfolg und die Testintelligenz (z.B. GRESSER-SPITZMUELLER, 1973; SCHORB & SCHMIDBAUER, 1973). Die Intelligenzleistung, wie sie die verwendeten Intelligenztests erfassen, ist, wie immer der Grad der Erblichkeit der die soziale Herkunft bestimmenden Statusfaktoren eingeschätzt wird, von der sozialen Herkunft abhängig. Dies gilt auch für die Leistungen in nicht-verbalen Intelligenztests (vgl. MERZ & STELZL, 1977, S. 82-83), die wegen ihrer vermeintlichen Unabhängigkeit von der sozialen Herkunft gerne als Einteilungshilfen herangezogen werden. Intelligenzleistungen ihrerseits sind wiederum eine wesentliche Bedingung für Schulleistungen. Auch spricht einiges dafür, dass die soziale Herkunft ausser dem (erblich und/oder umweltbedingten) Einfluss über die Intelligenzleistung eine direkte Wirkung auf die Schulleistung bzw. auf die Chance einer "höheren" Schulbildung bzw. Kurseinstufung ausübt (vgl. dazu die pfadanalytischen Ergebnisse amerikanischer Untersuchungen, z.B. SEWELL & SHAH, 1968; DUNCAN et al., 1972. Vergleichbare Untersuchungen in der BRD fehlen, doch dürften auch in der BRD die Ergebnisse nicht viel anders sein).

So verwundert es denn nicht, dass die Gruppierung der Schüler in beiden Leistungsfächern herkunfts- bzw. schichtenspezifisch verläuft (vgl. STARK, 1971; BERNHARDT et al., 1974, S. 83).

Das bei der ersten Zuweisung in die Kurse gegebene Bedingungsgefüge lässt sich - vereinfacht zwar, aber die wesentlichen Merkmale des Prozesses dennoch beachtend - so darstellen:

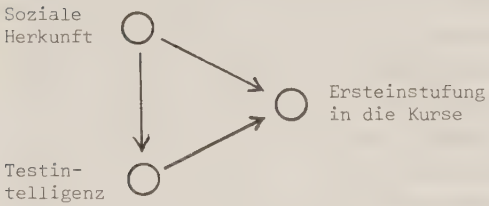

Abb. 1: Kausalstruktur zur Erklärung der Ersteinstufung in die Kurse.
Pfeile kennzeichnen die unterstellten Kausalbeziehungen. Die Richtungen dieser Kausalbeziehungen sind positiv.

Mit der Ersteinstufung wird also in der Regel soziale Auslese betrieben. Darüber, ob durch die organisatorische Massnahme dieser Art von Leistungsdifferenzierung allein die pädagogischen Zielvorstellungen erreicht werden, sind die Meinungen inzwischen geteilt (z.B. KEIM, 1973; WERRES, 1975).

Ein weiteres Kriterium für die Ersteinstufung in die Kurse ist die (eingeschätzte) Leistungs- bzw. Anstrengungsbereitschaft des Schülers (siehe die "Richtlinien für die Förderstufe", Absatz IV, Erlass des hessischen Kultusministers vom 14.3.1972). Am Ende der Förderstufenzeit, beim Uebergang in die Sekundarstufenkurse, sollen die Einstufungen unter anderem auch wieder von der Anstrengungsbereitschaft abhängig gemacht werden. Die Zuweisung zu homogenen Leistungsgruppen zu Beginn der Förderstufe bedeutet demnach auch eine niveauspezifische Vereinheitlichung in der Anstrengungsbereitschaft der Schüler, d.h. in den unteren Kursniveaus sitzen vornehmlich die Schüler mit der geringeren Anstrengungsbereitschaft. Nach Meinung der Befürwor-

ter der äusseren Leistungsdifferenzierung sollte es dadurch auch möglich sein, auf die Lernvoraussetzungen leistungsschwacher Schüler besser einzugehen mit dem Ziel, die Anstrengungsbereitschaft aller Schüler im Verlaufe der Förderstufe (5. und 6. Schuljahr) zu fördern. In der Gruppierung von Schülern mit ähnlicher Anstrengungsbereitschaft bzw. durch eine geringere Streuung der Leistungsvoraussetzungen der Schüler wird gerade für die schwachen Schüler die Möglichkeit gesehen, mangelnde Anstrengungsbereitschaft auszugleichen (vgl. dazu HECKHAUSEN, 1968, S. 212-212).

Geht man entsprechend des von der Förderstufe selbstgesetzten Ziels "Fördern statt Auslesen" davon aus, dass Mängel der Anstrengungsbereitschaft durch pädagogische Förderungsbemühungen behoben werden können, dann kann die Frage nach der Legitimation der Förderstufe für sozial gerechte Zuteilung von Berufs- und Lebenschancen nicht lauten, ob Leistungskurse zu Beginn der Förderstufe herkunftsspezifische Anstrengungsbereitschaften widerspiegeln, sondern ob sie die Zuordnung: Kursniveau - Anstrengungsbereitschaft bis zum Ende der Förderstufe durchbrechen oder nicht.

Eine Selektionsentscheidung könnte man dementsprechend zunächst so quantifizieren: Die unter anderem nach der Anstrengungsbereitschaft getroffene Erstzuordnung wird zu einer Selektionsentscheidung dann, wenn bei relativ konstanter, eindeutiger Personenzuordnung, also hier: bei relativ stabiler sowie relativ undifferenzierter, fachübergreifender Kurszuordnung die Korrelation zwischen Kurszuordnung und Anstrengungsbereitschaft am Ende der Förderstufenzeit so hoch ist, dass sie bei Kontrolle von Variablen pädagogischer Förderungsbemühungen (und selbst bei Kontrolle weiterer Variablen wie "soziale Herkunft", "Testintelligenz", usw.) nicht verschwindet. Dieser quantitativen Bestimmung entsprechend wäre es das Ziel der Förderstufe, präventiv möglichst so zu intervenieren, dass das Eintreffen einer

solchen Korrelation verhindert wird.

Da eine solche quantitative Bestimmtheit der (auf die Anstrengungsbereitschaft bezogenen) Zielsetzung aus den Zielsetzungen des herkömmlichen Schulsystems nicht abgeleitet werden kann, wäre ein diesbezüglicher Schulsystem-Vergleich in seinen Konsequenzen anfechtbar und willkürlich, d.h. je nach bildungspolitischem Standpunkt verschieden beurteilbar. Auf einen Intersystem-Vergleich muss daher verzichtet werden (zumindest solange, bis ein tragfähiger Konsens darüber hergestellt ist), und es kann im Sinne einer Produktevaluation nur überprüft werden, was die Förderstufe bzw. eine einzelne Förderstufen-Schule im Hinblick auf die von ihr selbst gesetzten Ziele leistet.

Von der vorgenannten Quantifikationsbestimmung ausgehend kann man nun einen Test der Zielsetzung "Fördern statt Auslesen", bezogen auf die Anstrengungsbereitschaft, dadurch weiter präzisieren, indem man a priori alternative Kausalmodelle zu formulieren versucht, die einem Entscheidungstest zugeführt werden können. Die folgenden Ueberlegungen dienen diesen Präzisierungsbemühungen.

Einige Autoren (z.B. HIELSCHER, 1970; LOSER, 1970) glauben zu erkennen, dass die Kurseinstufung der Schüler von vielen Lehrern als primär organisatorische Massnahme verstanden wird, die helfen soll, die Lehrtätigkeit zu erleichtern. Gestützt wird diese Annahme u.a. durch die Beobachtung, dass Lehrer versuchen, in unteren Kursniveaus ähnliche Schülerleistungen in gleicher Zeit zu erreichen, "wie sie der dafür ungewollt Masstäbe setzende A-Kurs braucht" (GEISSLER et al., 1969, S. 167). Verallgemeinernd wird deshalb auch behauptet, dass Lehrer sich durch die leistungshomogene Gruppierung zu einem undifferenzierten, d.h. didaktisch wenig flexiblen Lehrverhalten verleiten liessen, worunter gerade die C-Kurs-Schüler zu leiden hätten. So wird von GOLDBERG et al. (1969) darauf hingewiesen, dass der Wert der

Niveaugruppenbildung davon abhänge, wie er didaktisch genutzt werde. Kritiker einer nur organisatorisch verstandenen Leistungsdifferenzierung (z.B. NIEHL, 1975; HEPTNER, 1976) tendieren denn auch dazu, "schulische Förderung" mit "Förderung durch den Lehrer" gleichzusetzen und dafür zu plädieren, die Niveaugruppenbildung nicht als "Lehr-Erleichterung" anzusehen, sondern als eine Art "Uebersichts-Erleichterung" von Unterrichtsbemühungen. Fasst man nämlich die soziale Herkunft und die Kriterien der Erstzuweisung (insbesondere den bisherigen Schulerfolg und die Testintelligenz) als allerdings grobe Indikatoren für das Ausmass bisher erhaltener Förderung auf, dann sind mit einer solchen Niveaugruppierung zugleich die Schüler ausfindig gemacht, die der Förderung und Unterstützung durch den Lehrer in besonderem Masse bedürfen.

Es ist unbestritten, dass die Anstrengungsbereitschaft des Schülers auch vom didaktischen Geschick des Lehrers abhängt. Zu fragen ist aber, ob ausschliesslich durch den Lehrer kompensatorische Förderung der Anstrengungsbereitschaft erreicht werden kann. Manche empirische Arbeit belegt dies offenbar zur Genüge. Auf zwei Arbeiten, die diesen Schluss einer Kompensation durch den Lehrer nahelegen, soll hier besonders eingegangen werden. Die erste Arbeit stammt von BORCHERT & MASENDORF (1975), die zweite von SCHWARZER (1975).

BORCHERT & MASENDORF überprüften an einer Stichprobe von 11-12jährigen Schülern, in der Sonder-, Haupt- und Realschüler zusammengefasst waren, die kausale Hypothese, der Erziehungsstil der Mutter und der Lehrer vermittele den Einfluss der sozialen Herkunft eines Schülers auf seine Anstrengungsbereitschaft. Im folgenden Pfaddiagramm (s. Abb. 2) ist diese kausale Hypothese veranschaulicht. Bei dieser Darstellung lassen wir die von den Autoren miterhobene Variable "Schulangst" ausser acht. Diese Variable hat in ihrer Untersuchung offenbar einen der Variablen "Anstrengungsbereit-

schaft" analogen Kausalstatus.

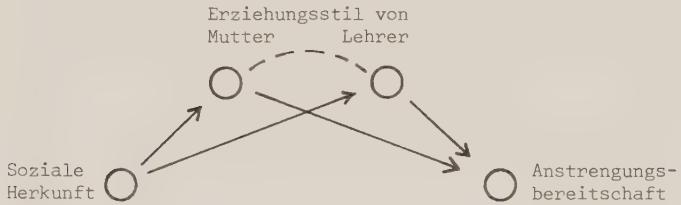

Abb. 2: Die von BORCHERT & MASENDORF (1975) postulierten Kausalbeziehungen zur Erklärung der Anstrengungsbereitschaft (----: diese Variablenbeziehung wurde von den Autoren kausal nicht spezifiziert)

Die Anstrengungsbereitschaft wurde durch den Anstrengungsvermeidungstest (AVT) von ROLLETT & BARTRAM (1972) erfasst und der Erziehungsstil durch eigens (von BORCHERT) konstruierte Skalen. Die Items dieser Skalen sind teilweise den Marburger Erziehungsstil-Skalen von HERRMANN et al. (1968) und STAPF et al. (1972) entnommen. Mutter- und Lehrerskala stimmen in Anzahl und inhaltlichen Formulierungen der Items überein. Jede Skala erfasst in einer einzigen Dimension das positive Bekräftigungsverhalten (Unterstützung) und das negative Bekräftigungsverhalten (Strenge) zugleich. BORCHERT & MASENDORF messen also beim Test ihrer Hypothese den Erziehungsstil allgemein, ohne das Ausmass der von HERRMANN & STAPF als faktoriell unabhängig verstandenen Bekräftigungskomponenten "Unterstützung" und "Strenge" im einzelnen aufzuschlüsseln, d.h. ein niedriger Skalenwert in der Mutter- oder Lehrerversion bedeutet für sie nicht nur niedrige Bekräftigung, sondern auch geringe Unterstützung bzw. geringes Belohnungsverhalten (S. 144 bzw. S. 145).

BORCHERT & MASENDORFs Aussagen zufolge stützen die Untersuchungsergebnisse die in Abb. 2 dargestellte Kausalhypothese (s. S. 137 und S. 144). Allerdings hält ihre Datenanalyse einer kritischen Ueberprüfung nicht stand. So ist zum Beispiel nicht ganz klar, was sie unter "Vermittlung eines Einflusses" (s. S. 138) verstehen. Wenn man im Kontext von Kausalhypothesen von "Vermittlung" spricht, ist üblicherweise eine einfache kausale Kette damit gemeint, nämlich: "soziale Herkunft" ⟶ "Erziehungsstil" ⟶ "Anstrengungsbereitschaft". Auf diesen Sachverhalt erscheint ihre statistische Behandlung des Problems jedoch nicht zugeschnitten (In einem separaten Aufsatz haben wir diese Behauptung näher begründet. Dieser Aufsatz kann angefordert werden). Wir vermuten daher, dass BORCHERT & MASENDORF mit ihrer "Vermittlungshypothese" gar nicht ausschliesslich eine einfache Kausalkette meinen, sondern eine Kausalstruktur, die neben einer indirekten auch eine direkte Kausalbeziehung von der sozialen Herkunft auf die Anstrengungsbereitschaft miteinbezieht.

SCHWARZER (1975, etwas kürzer auch in 1976 a, b. Wir zitieren im folgenden aus den leichter zugänglichen Arbeiten von 1976) untersuchte das Urteilsverhalten von Lehrern gegenüber den ihnen anvertrauten Erstklässlern. Die Lehrer sollten die Persönlichkeit eines jeden Schülers beurteilen, und zwar mit Hilfe von Eigenschaftsbegriffen, die einer Untersuchung HOFERs (1969) entnommen worden waren. Zum Zeitpunkt ihrer Beurteilung hatten die Schüler bereits ein halbes Schuljahr hinter sich. Die Autorin fand, dass die Lehrer dazu neigen, Unterschichtkinder eher als uninteressiert, dumm, träge und unsympathisch zu beurteilen: "Das Lehrerurteil ist sozialschichtabhängig" (1976 a, S. 312).

Dieser Befund stimmt mit Ergebnissen anderer empirischer Arbeiten überein, die darauf hinweisen, dass das Wissen um das Herkunftsmilieu des Kindes das Lehrerurteil im Hinblick auf

den Uebergang auf weiterführende Schulen beeinflusst (z.B. PREUSS 1970; STEINKAMP, 1972). Man muss daher davon ausgehen, dass diese Leistungsbewertung der Lehrer mit zur schichtspezifischen Auslese bei der Ersteinstufung in die Förderstufenkurse beiträgt.

Weiterhin ermittelte SCHWARZER,

> "dass das Urteilsverhalten von Lehrern ungefähr gleichermassen von der Sozialschichtzugehörigkeit und von der sprachlichen Intelligenz der Kinder (erfasst mit einem Wortschatztest, dem "Peabody Picture Vocabulary Test", der Verf.) abhängt. Die sprachliche Intelligenz ist ihrerseits wieder teilweise vom sozio-ökonomischen Status abhängig" (1976 a, S. 312)

Die kausalen Beziehungen "zwischen dem Urteilsverhalten und seinen wichtigsten Einflussgrössen" stellt SCHWARZER (s. auch 1976 a, S. 312) so dar:

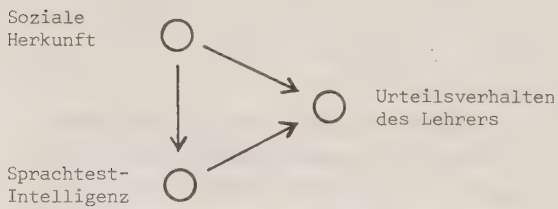

Abb. 3: Die von SCHWARZER (1975) postulierten Kausalbeziehungen zur Erklärung des Lehrer-Urteilsverhaltens

Zunächst einmal ungeachtet dessen, dass Einzelbefunde verschiedener Studien keinen hinreichenden Beleg dafür liefern, eine umfassendere Theorie im nachhinein kausal zu interpretieren: Zieht man Untersuchungsergebnisse von WEINER & KUKLA (1970) und von BROPHY & GOOD (1970) mit in Betracht, dann liegt es nahe, beide hier besprochenen Studien in ihren Hypothesen widerspruchsfrei wie folgt (s. Abb. 4) zu kombinieren:

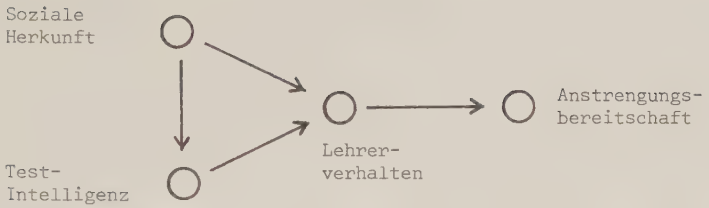

Abb. 4: Kombination der Kausalstrukturen von Abb. 2 und Abb. 3 zur Erklärung der Anstrengungsbereitschaft

Die eben genannten Autoren stellten fest, mit der Ursachenerklärung des Lehrers für Schülerverhalten gehe im allgemeinen eine gleichgerichtete affektive Reaktion einher: Der Lehrer wende sich denjenigen Kindern bevorzugt (akzeptierend und unterstützend) zu, von denen er eine positive Meinung und Leistungserwartung habe, in unserem Fall den Kindern mit "guter" sozialer Herkunft und Testintelligenz.

Die Darstellung in Abb. 4 zeigt die Schlüsselstellung besonders deutlich, die dem Lehrer in der Entwicklung der Anstrengungsbereitschaft eingeräumt wird. Es scheint von daher sehr wichtig zu sein, Fördermassnahmen in der Schule vorwiegend über die Person des Lehrers zu versuchen. BORCHERT & MASENDORF deuten dies an, wenn sie von einem "kompensatorischen Bekräftigungsverhalten" (1975, S. 144/45) durch den Lehrer sprechen. Auch der von SCHWARZER gemachte Förderungsvorschlag liegt in dieser Richtung, wenn sie für eine "kompensatorische Beurteilung" durch den Lehrer plädiert:

> "Wenn nachgewiesen ist, dass der Lehrer unwissenderweise sozialschichtabhängig urteilt, kann man eine so gravierende Benachteiligung nicht einfach hinnehmen. Es wird vorgeschlagen, in Analogie zur kompensatorischen Erziehung eine "kompensatorische Beurteilung" einzuführen. Praktisch bedeutet das die Einführung eines "Unterschicht-Bonus", der der benachteiligten Gruppe bei der Beurteilung implizit oder auch offen gutgeschrieben würde. Dabei ist noch unklar, wie so etwas konkret geschehen könnte. .... Der Schüler soll sich wohlfühlen, ein positives Selbstbild entwickeln, den Unterricht aktiv steuern helfen und dabei Lernbereitschaft und Interesse entwickeln. Auf Seiten des Schülers erfordert das einen grösseren Entscheidungsspielraum, auf Seiten des Lehrers ein pädagogisch-diagnostisches Bewusstsein und eine höhere therapeutisch-didaktische Kompetenz" (1976 a, S. 313)

Abschliessend ist zu den beiden hier ausführlicher referierten Arbeiten zu sagen: Sie unterstellen beide, wenn auch explizit nicht so formuliert, dass kompensatorisches Lehrerverhalten vorliegt, wenn der (korrelative) Zusammenhang zwischen der sozialen Herkunft der Schüler und dem (erlebten Unterstützungs-) Verhalten der Lehrer verschwindet.

Verständlicherweise erhofft man sich von einer verbesserten sozialpsychologischen Lehrerausbildung Hilfe bei der Bewältigung motivationspsychologischer Schülerprobleme. Der Lehrer sollte in die Lage versetzt werden, von seinen (im Laufe der eigenen Sozialisation erworbenen) sozialen Stereotypen und Vorurteilen abzusehen und seine Schüler unabhängig von ihrer sozialen Herkunft zu beurteilen und zu behandeln. Es liegt nahe, solche Vorstellungen auch in die aktuelle Förderstufen-Diskussion (oder auch allgemeiner: in die Gesamtschul-Diskussion) einzubringen, zumal sich angesichts unbefriedigender bzw. uneinheitlicher Reformresultate die Stimmen mehren, Innovationen in der Schule seien entscheidend nur über die Person des Lehrers durchzusetzen (z.B. WASNA, 1973; NIEHL, 1975).

Zusammenfassend: In einer "Zusammenlegung" der drei Modelle von Abbildung 1,2 und 3 (Abb. 2 inklusive der erweiterten Definition von Vermittlung) können die vorgenannten Kausalhypothesen widerspruchsfrei zu einem komplexen Kausalmodell zusammengefasst werden. Auf diese Weise ergibt sich eine Kausalstruktur, für die folgendes Merkmal charakteristisch ist: Die kausale Beziehung von der sozialen Herkunft und von der Testintelligenz auf die Anstrengungsbereitschaft wird nur über pädagogische Förderungsbemühungen (der Eltern und Lehrer) vermittelt, mit anderen Worten: Es besteht weder eine direkte noch eine indirekte kausale Beziehung von der Kurszugehörigkeit zur Anstrengungsbereitschaft, eine Kausalstruktur, die man vereinfachend auch bezeichnen könnte als "keine Vermittlung durch die Kurszugehörigkeit". Die folgenden Ueberlegungen

dagegen lassen eine alternative Kausalstruktur plausibel erscheinen.

Die Tatsache, dass Schüler die Leistungen vollbringen und Lehrer diese Leistungen beurteilen, mag mitunter dazu verleiten, Schulleistungsschwierigkeiten in erster Linie als ein individuelles Problem konkreter Personen zu verstehen. Zumindest sind die Untersuchungen von BORCHERT & MASENDORF und von SCHWARZER so angelegt, dass aus ihren Ergebnissen Rechtfertigungen für Veränderungen im Schulsystem praktisch nur innerhalb der Person des Schülers bzw. Lehrers gefunden werden können. Es ist aber zu beachten, dass ein statistisches Ergebnis, welches der kausalen Hypothese nicht widerspricht, diese nicht unbedingt auch bestätigt. Denn immerhin ist eine alternative Erklärung denkbar, die einen fehlenden direkten Einfluss von der sozialen Herkunft (bzw. der Testintelligenz) auf das Lehrer- und Schüler-Verhalten, also eine Scheinbeziehung behauptet, aber ebenso gut zu den Daten passt. Eine solche Alternativerklärung sehen wir, allgemein formuliert, darin, dass institutionelle Bedingungen der Schule den Zusammenhang zwischen der sozialen Herkunft der Schüler und dem Lehrer- sowie Schüler-Verhalten vermitteln.

Im folgenden versuchen wir diese mögliche Alternative zu erklären, zunächst in Bezug auf die Person des Lehrers, dann in Bezug auf die Person des Schülers. Diese Alternativerklärung werden wir später - dies sei hier schon vorweggenommen - zusammenfassend als Hypothese von der "entscheidenden Vermittlung durch die Kurszugehörigkeit" bezeichnen.

1.2. Die Institution "Schule" als Bedingung für die Unterstützung durch den Lehrer und für die Anstrengungsbereitschaft des Schülers

Es steht ausser Frage, dass die Einstellungen der Lehrer zur Förderstufe, ihre Bereitschaft, in der neuen Schulform mitzuarbeiten, eine wichtige Bedingung für den wie immer definierten Erfolg der Förderstufe sind. So legen z.B. die Ergebnisse von ZEIHER (1973) die Vermutung nahe, dass die ablehnende oder zumindest ambivalente Einstellung besonders der (ehemaligen) Gymnasiallehrer gegenüber Schulreformen sich als Hindernis für den Erfolg der Förderstufe erweisen könnte. Eigene informelle Befragungen ergaben zudem, dass die meisten Förderstufenlehrer am liebsten in den höheren Kursen unterrichten (vgl. auch HURRELMANN, 1971, S. 58). Alle diese Beobachtungen stützen anscheinend die Auffassung, dass "jede wirksame Schulreform steht und fällt mit einer entsprechenden Reform der Lehrerausbildung" (KRAMP, 1973, S. 116).

Vergegenwärtigt man sich die Ergebnisse einer 1974 im Kreis Wetzlar durchgeführten Infra-Test-Umfrage, so kann man sich lebhaft vorstellen, dass diese Auffassung von den Förderstufenlehrern selber als Tadel an den persönlichen Leistungen empfunden bzw. als Versuch gewertet werden muss, der Schulreform, genauer: der in der Förderstufe praktizierten Form der Leistungsdifferenzierung ein Alibi zu verschaffen. Die Mehrzahl der 250 von Infra-Test befragten Förderstufenlehrer nämlich führt die geringere Unterstützung der C-Kurs-Schüler auf die Ueberlastung der Lehrer mit organisatorischen Aufgaben zurück. Gerade dies mag erklären, warum die meisten Förderstufenlehrer den Unterricht in den höheren Kursen bevorzugen. In Zeitnot beschränkt man sich möglicherweise auf ein Aktivieren nur der leistungsstärkeren Schüler. Erfolge stellen sich bei diesen müheloser und auch augenfälliger ein und das Gefühl des Vor-

wärtskommens bekräftigt dann unversehens und unbeabsichtigt die Tendenz, sich um die Schüler höherer Kursniveaus besonders zu bemühen. Für 87% der in der Infra-Test-Untersuchung befragten Lehrer stellt es darum auch ein Problem dar, dass "die Schulreform auf dem Rücken der Lehrer ausgetragen wird (stärkere Arbeitsüberlastung, usw.)".

Dass Schulreform nur dann sinnvoll ist, wenn sie nicht am Lehrer vorbei, sondern unter seiner Mitarbeit durchgeführt wird, bedarf wohl keiner näheren Begründung. Und sicher richtig ist auch die Auffassung, dass entlastete Lehrer bessere Lehrer sind. Zu bezweifeln ist aber, ob diejenigen recht haben, die aus den Infra-Test-Ergebnissen glauben ableiten zu müssen, bei Abbau dieser Arbeitsüberlastung sei der Lehrer in die Lage versetzt, C-Kurs-Schüler zu fördern.

Eine andere Erklärung für das beobachtete kursspezifische Verhalten der Lehrer sehen wir darin, dass die Ersteinstufungen in die Kurse zwar von einem wie immer verzerrenden Lehrer(vor)urteil mitbestimmt sein mögen, dass aber die Kurse, wenn sie erst einmal eingerichtet sind, von den Lehrern zugleich auch als Indikatoren der allgemeinen schulischen Leistungsfähigkeit gewertet werden (müssen). Von den Kurseinstufungen der Schüler wird auf die schulische Leistungsfähigkeit geschlossen, so vermuten wir, weil man sich das Zustandekommen von Schulleistungen primär mit diesem Kausalfaktor erklärt. Ueber allgemeine Leistungserwartungen vermittelt können diese Einschätzungen der Leistungsfähigkeit dann das Lehrerverhalten bestimmen. Die schulische Leistungsfähigkeit kommt für den Lehrer zur Erklärung der Kurszugehörigkeit unserer Meinung nach deshalb als (hauptsächlicher) Kausalfaktor in Betracht, weil dieser eine allgemein übliche und auch verbindliche Beurteilungskategorie der Institution "Schule" darstellt. Die Schule selbst ist bei der Vergabe von Qualifikationsnachweisen an Normen gebunden, die, wenn auch vermittelt, gesellschaftliche Ansprüche enthal-

ten. Eine solche Einschätzung der Kurszugehörigkeit wäre demnach nicht eine willkürliche, private Konstruktion des Lehrers, sondern hinge mit seiner Berufsrolle zusammen.

Dass in der Förderstufe die allgemeine schulische Leistungsfähigkeit (vorwiegend) nach den Kurseinstufungen eingeschätzt wird, könnten indirekt folgende Beobachtungen nahelegen. Im Laufe der Förderstufenzeit erweist sich die Kurszugehörigkeit der Schüler als ziemlich stabil (vgl. z.B. GEISSLER et al., 1967; STARK, 1971). Sieht man es als Ziel der Förderstufenversuche an, anfangs leistungsschwache Schüler auch hinsichtlich ihrer Schulleistungen zu fördern, müssten sich niedrigere Stabilitäten ergeben. Oft ist auch eine Polarisierung des Schülerjahrgangs in eine leistungsstarke und in eine leistungsschwache Gruppe zu beobachten: In einem als "setting" (d.h. fächerspezifisch) angesetzten Kurssystem werden die unteren Leistungsgruppen im Laufe der Förderstufenzeit zunehmend fächerübergreifend, also "streaming"-ähnlich organisiert.

Angesichts dieser fächerunabhängigen Segmentierung der Leistungsgruppen - man könnte hier etwas vergröbernd statt von "den" Kurszugehörigkeiten zusammenfassend auch von "der" Kurszugehörigkeit sprechen - kann man sich fragen, ob im Laufe der Förderstufenzeit nicht eine Betrachtungsweise der Lehrer (und auch Eltern) durchschlägt, die das Kurssystem im Sinne einer Drei- oder Zwei-Gliederung des hergebrachten Schulsystems interpretiert. Die durch Elternwunsch und Lehrerurteil zustandegekommenen Ersteinstufungen scheinen hier eher einen sich selbst bestätigenden als diagnostischen Charakter zu haben: Es liegt die Vermutung nahe, dass in den Kurseinstufungen gegen Ende der Förderstufe so etwas wie ein gemeinsamer Faktor der allgemeinen Leistungserwartung, eine pauschale Erwartungshaltung bezüglich der schulischen Leistungsfähigkeit zum Ausdruck kommt. Mit anderen Worten, es ist anzunehmen, dass die (Fremd-)Einschätzung der schulischen Leistungsfähigkeit kursabhängig ist.

Man kann davon ausgehen, dass Lehrer an der Förderstufe sich vor allem am gymnasialen Standard der Leistungen orientieren, da dies für die Oeffentlichkeit das entscheidende Merkmal für die Qualität der neuen Schulform ist. Es liegt daher nahe, dass der Lehrer die Leistungsentwicklung derjenigen Schüler besonders beobachtet und fördert, bei denen er die Eignung für "höhere" Ausbildungsgänge vermutet, und diese Vermutungen werden von den Kurseinstufungen bestimmt.

Die empirische Berücksichtigung der Annahme, dass die Kurseinstufungen von den Lehrern zugleich auch als Indikatoren der allgemeinen schulischen Leistungsfähigkeit gewertet werden (müssen), kann dazu führen, dass der von BORCHERT & MASENDORF und von SCHWARZER postulierte Zusammenhang zwischen der sozialen Herkunft der Schüler und dem Lehrerverhalten verschwindet, ohne dass von einer kompensatorischen Förderung durch den Lehrer die Rede sein kann.

Im folgenden versuchen wir die mögliche Alternative, die Hypothese von der entscheidenden Vermittlung durch die Kurszugehörigkeit, nun in Bezug auf die Person des Schülers zu erklären. Die Vorstellung einer kompensatorischen Förderung durch den Lehrer bzw. die Forderung, dass mit der Einrichtung der Förderstufe zugleich auch das Förderverhalten der Lehrer reformiert werden müsste, setzt die Bedingung voraus, dass die Kurszugehörigkeit nur auf das Förderverhalten direkt wirkt, so dass der Effekt der Kurszugehörigkeit auf die Anstrengungsbereitschaft der Schüler vollkommen über das Förderverhalten vermittelt wird. Mit anderen Worten, es muss der zwischen der Kurszugehörigkeit und der Anstrengungsbereitschaft der Schüler beobachtete Zusammenhang bei Kontrolle des Lehrerverhaltens verschwinden.

Man muss aber davon ausgehen, dass mit den Kurszugehörigkeiten allgemeingültige, traditionelle Erwartungen und Selbstverständlichkeiten verknüpft sind, denen sich die Schüler

auch bei Kontrolle des Lehrerverhaltens, z.B. auch bei engagiertestem Einsatz der Lehrer für die C-Kurs-Schüler, nicht entziehen können, d.h. institutionelle Bedingungen der Schule wirken sich direkt auf Einstellung und Verhalten von Lehrern und Schülern aus. Denn es gibt deutliche Hinweise dafür, dass auch die Schüler ihre Kurszugehörigkeiten als Indikatoren der schulischen Leistungsfähigkeit ansehen. "Wir sind dumm, sonst wären wir nicht im C-Kurs", so lautete nach Auskunft eines Lehrers (HUESKE, 1973, S. 346) das Urteil seiner C-Kurs-Schüler über sich selbst. Auch Ergebnisse von FEND et al. (1975, S. 14) stützen diese Annahme der kursspezifischen Begabungs(selbst-)einschätzung. Dementsprechend ist in der Alternativerklärung auch die Vermutung mitenthalten, dass institutionelle Bedingungen der Schule den Einfluss von der sozialen Herkunft auf die Anstrengungsbereitschaft direkt weiter vermitteln.

Institutionellen Bedingungen mangelnder Anstrengungsbereitschaft wurde in der empirischen Forschung bisher relativ wenig Aufmerksamkeit entgegengebracht. Das ist allerdings nicht erstaunlich, denn Institutionen sind für den Menschen so etwas wie eine zweite Natur und nehmen daher sehr rasch den Charakter von Selbstverständlichkeiten an. Aber gerade weil der institutionelle Bedingungskomplex so wenig offensichtlich ist, ist er wichtig, und daher erfordert die Hypothese einer Vermittlung durch institutionelle Bedingungen eine genauere Erklärung. Eine solche Erklärung könnte im kombinierten Auftreten dieser beiden institutionellen Bedingungen liegen: der Ueberbewertung der Schulleistungen im Rechnen und in den sprachlichen Fächern und der gruppenbezogenen Beurteilung dieser Leistungen.

Im herkömmlichen Schulsystem der BRD bezieht sich die Beurteilung der Leistungen in Form von guten Noten, Lob, usw. durch den Lehrer primär auf "richtige", "gute" oder auf vollständige Leistungen, d.h. ein Schüler, der noch zu keiner Lösung gekommen ist, erfährt vom Lehrer gewöhnlich keine positive Beurteilung. Es kommt hinzu, dass der Schüler nicht individuell danach

beurteilt wird, ob er hinsichtlich seiner Leistungsfähigkeit oder seinen bisherigen Leistungen besser oder schlechter gearbeitet hat. Vielmehr wird er regelmässig mit den Schülern seiner Gruppe verglichen. Wird aber die Leistungsverteilung einer Gruppe als Beurteilungsmasstab genommen, dann gibt es unabhängig von den Lernfortschritten zwangsläufig eine Subgruppe schlechter Schüler. Legt man diesen Masstab auch noch in institutionell geregelten Zeitabständen an, dann stellt die Leistungsbeurteilung eine Unterbrechung des Lernvorganges für die Schüler dar, die noch nicht zu einer richtigen oder vollständigen Lösung gekommen sind. Wenn sich Lehrer an diesen Masstäben der Beurteilung orientieren, entsprechen sie durchaus den in unserem Schulsystem gängigen Erfolgs- und Auslesekriterien.

Auf dem Hintergrund einer solchen Lernkontrolle wird aber auch die Entwicklung schulischer Anstrengungsvermeidung verständlich: Ein in der Leistungshierarchie auch nach Wochen der Anstrengung noch unter dem Gruppendurchschnitt rangierender Schüler muss seine verfestigt niedrige Einstufung als Scheitern seiner Anstrengungsbemühungen deuten, selbst wenn seine Fortschritte in Richtung auf ein festgelegtes Leistungsziel hin beträchtlich waren. Die Nichtbeachtung sowie Unterbrechung von Lernanstrengungen eines als leistungsschwach eingestuften Schülers kann als Bestrafung lernerischer Aktivitäten aufgefasst werden. Erlebt der Schüler, dass Fehler zu machen für ihn mit negativen sozialen Konsequenzen verbunden ist, dann versucht er der Strafdrohung aus dem Wege zu gehen: Er meidet, wenn er kann, schulische Leistungssituationen.

Die Schulleistung hat insofern soziale Konsequenzen, als die soziale Anerkennung durch Lehrer (aber auch durch Eltern und Mitschüler) davon abhängig gemacht wird. Schlechte Schulleistungen mindern die erlebte soziale Anerkennung. In einer Untersuchung an 10-12jährigen Schülern fanden DAVIDSON & LANG (1960), dass die leistungsschwächeren Schüler die Gefühle und

Einstellungen ihrer Lehrer ihnen gegenüber häufiger negativ einschätzten als die leistungsstärkeren Schüler. Offenbar wird die Schulleistung zum zentralen Kriterium für die (erlebte) soziale Anerkennung des Schülers durch den Lehrer. Ein für den leistungsschwächeren Schüler verhängnisvoller Zirkel von Leistung und Anerkennung könnte dann daraus resultieren, dass der Lehrer erst auf gute oder vollständige Leistungen hin positiv bekräftigend reagiert. Denn mangelnde Anerkennung und Unterstützung durch den Lehrer untergräbt das so dringend benötigte Vertrauen in die eigene Leistungsfähigkeit und damit eine unabdingbare Voraussetzung zur Anstrengungsbereitschaft.

Von der gruppenbezogenen Leistungsbeurteilung ist die Lehrplanproblematik (z.B. WILHELM, 1967, S. 208) nicht zu trennen. Schulische Leistungsanforderungen beschränken sich auf bestimmte Unterrichtsinhalte. Eine institutionell verankerte Bedingung mangelnder Anstrengungsbereitschaft liegt möglicherweise in der Tatsache, dass vorwiegend die Leistungen in sprachlichen Fächern und in Rechnen über den weiterführenden Bildungsgang entscheiden.

Die Vorrangstellung dieser Fächer zeigt sich unter anderem darin, dass Lehrer dort die strengsten Noten erteilen (HOPP & LIENERT, 1965; FERDINAND & KIWITZ, 1971). Dieser Befund lässt sich auch damit erklären, dass Lehrer die Leistungen in Rechnen und in den Sprachen als zulängliche Indikatoren der allgemeinen schulischen Leistungsfähigkeit betrachten, d.h. die Leistungen in den anderen Fächern werden eher als nebensächlich angesehen. Doch Lehrer müssen sich notwendig so verhalten; denn Gesellschaft und Oeffentlichkeit erwarten von ihrem Schulsystem (auch) die funktionsgerechte Erledigung der Auslese (Selektion) und sozialen Zuteilung (Allokation), und nach allgemeinem Dafürhalten sind nun einmal Rechnen und Sprachen die für die Vergabe von Bildungs- und Berufschancen entscheidenden Fächer.

In dieses Bild fügen sich auch die folgenden Befunde ein. Nach BECKER (1968) glauben Lehrer immer dann ihre Aufgabe in der Schule gut zu verrichten, wenn sie durch den Unterricht die intellektuellen Fähigkeiten und Kenntnisse der Schüler deutlich erweitern. Stoff- und Wissensvermittlung stehen im Mittelpunkt der Unterrichtsbemühungen. Doch die Schüler selbst, ihre Gedanken und Gefühle, gehören nicht zu den Themen des Unterrichts. Nach HOFERs (1969) Ergebnissen wird die Persönlichkeit des Schülers vorwiegend unter "schulisch relevanten", d.h. hier: leistungsorientierten Merkmalen bewertet. Andere Merkmale spielen bei der allgemeinen Persönlichkeitsbeurteilung des Schülers durch den Lehrer eine eher untergeordnete Rolle:

> "Der Faktor der menschlichen Zuneigung, der in anderen Untersuchungen zur impliziten Persönlichkeitstheorie hervorsticht, ist hier nur von untergeordneter Bedeutung gegenüber dem leistungsorientierten Aspekt in der sozialen Wahrnehmung des Lehrers vom Schüler" (HOFER, 1969, S. 87)

Diese vorwiegend am Masstab der schulischen Leistung orientierte Sichtweise der Lehrer erklärt möglicherweise auch, warum die Schülereigenschaft "un-/interessiert" im Eindruck der Lehrer so zentral ist ("zentral" heisst hier: In der erwähnten Untersuchung HOFERs hatte die Eigenschaft "un-/interessiert" substantielle Ladungen auf den leistungsorientierten Faktoren, S. 47-48). Viele Lehrer wissen recht wenig über die Interessen und Lieblingsbeschäftigungen der ihnen anvertrauten leistungsschwachen Schüler (z.B. OSTERLOH & ROLAND, 1970). Gerade deshalb rückt möglicherweise auch der Beurteilungsgesichtspunkt "un-/interessiert" für den Lehrer so sehr in den Mittelpunkt. Denn was die Schule als (Lern-)Interesse definiert, bezieht sich auf ausgewählte Gegenstandsbereiche, nämlich vorwiegend auf solche, die intellektuell-abstrahierende Fähigkeiten verlangen, und damit lässt sich weniger gut an die Erfahrungen von Schülern anknüpfen, die von ihrer vor- und ausserschulischen Lerngeschichte her weniger Gelegenheit hatten, Interessen dafür zu entwickeln.

Man kann daher annehmen, dass über dieses Lerninteresse vermittelt Art und Zusammenstellung der Unterrichtsfächer die Anstrengungsbereitschaft beeinflussen. Wenn der Schüler kein Interesse an bestimmten Fächern mitbringt, werden die damit zusammenhängenden Leistungssituationen bei einigen Misserfolgen und bei einem wenig unterstützenden Lehrerverhalten einfach nicht mehr aufgesucht.

Bisherige Förderstufen-Erfahrungen weisen darauf hin, dass sich die hier zunächst auf die herkömmliche Schule bezogene Bedingungskombination - Ueberbewertung von Rechnen und Sprachen und gruppenbezogene Leistungsbeurteilung - mit ihren vermuteten Auswirkungen auf Lehrer und Schüler auch in der Förderstufe wiederfindet:

- Bei der gruppenbezogenen Leistungsbeurteilung wird der gesamte Schuljahrgang, das Kurssystem zum Bezugspunkt der Leistungsbeurteilung gemacht (vgl. z.B. BERNHARDT et al., 1974; FEND et al., 1975), d.h. die Subgruppe "schlechter" Schüler besteht nicht aus den jeweils "schlechten" Schülern innerhalb jeder der drei Kursniveaugruppen, sondern aus der C-Kurs-Gruppe überhaupt.
- Durch die stabile Kurseinstufung in Fächern, die in der BRD traditionsgemäss für den Schulabschluss massgebliche Bedeutung haben, erhalten die sog. Leistungsfächer einen höheren Stellenwert als die Fächer, die in Kerngruppen unterrichtet werden. Die Zugehörigkeit zu einer hohen Niveaugruppe wird bereits in der Förderstufe als Sprungbrett zu "höherer" Berufsbildung, zu höherem Sozialprestige angesehen. Die Schüler selber neigen dazu, den Unterricht in den Kernfächern gegenüber dem Kursunterricht als zweitrangig oder weniger wichtig zu werten (siehe z.B. die Erfahrungsberichte in KEIM, 1973). Die höheren Kursniveaus, mit "höherem" Qualifikationsniveau gleichgesetzt, erhalten den Rang von Vorstufen zum Gymnasium. Ihnen steht das C-Kurs-Niveau gegenüber, dem eine "volkstüm-

liche" Bildung (Hauptschule) bzw. der berufsvorbereitende Schulabschluss zugeordnet werden (BERNHARDT et al., 1974, S. 163-165). Da an die Schulabschlüsse unterschiedliche Berechtigungen zu bestimmten Berufslaufbahnen gekoppelt sind, erhalten die in den Förderstufen sich bildenden starren Leistungsdifferenzierungen Selektionscharakter im Sinne schwer korrigierbarer Schullaufbahnlenkungen. Der Schulerfolg wird somit institutionell in den Kurszugehörigkeiten verankert, d.h. dass aus der Sicht der stabil quasi-streaming organisierten Förderstufe vor allem derjenige ein schlechter Schüler ist, der den schulischen Anforderungen speziell in Mathematik und Sprachen nicht genügt.

1.3. Hypothesen dieser Untersuchung, noch nicht prüfbar und noch unvollständig

Aus allen diesen Ueberlegungen leiten wir die Erwartung ab, dass bei Berücksichtigung der Kurszugehörigkeit die Lehrerunterstützung und die davon abhängige Anstrengungsbereitschaft nicht einfach und direkt auf die soziale Herkunft und die Testintelligenz des Schülers zurückgeführt werden können. Wenn die Kurszugehörigkeiten nach institutioneller Auslegung als zulänglich bzw. repräsentativ für die schulische Leistungsfähigkeit gelten, dann dürften die Lehrerunterstützung und die Anstrengungsbereitschaft nur von diesen Kurszugehörigkeiten abhängen. Dann aber ist nicht die Lehrerunterstützung, sondern die Kurszugehörigkeit in Mathematik und Sprachen die entscheidende, den Einfluss von der sozialen Herkunft und von der Testintelligenz auf die Anstrengungsbereitschaft vermittelnde Variable. Fassen wir diese kausale Hypothese anschaulich in einem Pfaddiagramm zusammen (s. Abb. 5):

Abb. 5: Kausalstruktur zur Erklärung der Anstrengungsbereitschaft: die schulische Leistungsfähigkeit bzw. Kurszugehörigkeit in Mathematik und Sprachen als entscheidende vermittelnde Variable (sämtliche Kausalbeziehungen hier als positiv gerichtet postuliert)

Die in Abb. 5 dargestellte Hypothese von der entscheidenden Vermittlung durch die Kurszugehörigkeit kann man als mit der Hypothese SCHWARZERs nicht verträglich ansehen, sofern man unterstellt, dass Urteils- und Unterstützungsverhalten des Lehrers nicht wesentlich voneinander abweichen. Ebenfalls steht sie zu der Vermittlungshypothese BORCHERT & MASENDORFs in Widerspruch, d.h. wir erwarten, dass deren Vermittlungshypothese nicht so allgemeingültig ist, wie sie offenbar formuliert wurde. Dies gilt unabhängig davon, ob wir der Vermittlungshypothese den oben geschilderten weiteren oder engeren Sinn unterlegen.

Bevor wir weitergehen, erscheinen an dieser Stelle noch zwei Bemerkungen zur Begriffsverwendung und Definition angebracht.

- Wie bis hierher ersichtlich, wird die Modellbildung der Arbeit zugespitzt auf die Alternative "entscheidende Vermittlung durch die Kurszugehörigkeit" versus "keine Vermittlung durch die Kurszugehörigkeit". Um den Gegensatz zu pointieren, kann man zur zweiten Möglichkeit ("keine Vermittlung durch die Kurszugehörigkeit") auch sagen: "keine entscheidende Vermittlung durch die Kurszugehörigkeit". In der hiermit gekennzeichneten Kausalstruktur ist die Vermittlungshypothese BORCHERT & MASENDORFs widerspruchsfrei enthalten. Dementsprechend steht der Hypothese von der entscheidenden Vermittlung durch die Kurszugehörigkeit als äquivalente kausalstrukturelle Alternative zugleich auch die Vermittlungshypothese BORCHERT & MASENDORFs gegenüber. Verwirrungen aufgrund des gemeinsamen Teilbegriffs "Vermittlung" dürften nicht entstehen, wenn man beachtet, dass im folgenden für die zwei verschiedenen Arten von Vermittlung konsistent zwei verschiedene Kennzeichnungen verwendet werden, nämlich zum einen: "Vermittlungshypothese" und zum anderen: "Hypothese von der entscheidenden Vermittlung durch die Kurszugehörigkeit". Wenn in den weiteren Ausführungen einfach nur von "Vermittlungshypothese" die Rede ist, ist die Vermittlungshypothese BORCHERT & MASENDORFs damit gemeint, nicht die Hypothese von der entscheidenden Vermittlung durch die Kurszugehörigkeit, oder anders herum gesagt: Wenn die Hypothese von der entscheidenden Vermittlung durch die Kurszugehörigkeit gemeint ist, wird nie nur von "Vermittlungshypothese" gesprochen.

- Auf ausführlichere Definitionen der Variablen "Lehrerunterstützung" und "schulische Anstrengungsbereitschaft" hatten wir bislang verzichtet. Denn da es in dieser Arbeit auch darum geht, die von BORCHERT & MASENDORF vertretene Kausalhypothese zu testen, erscheint es angebracht, zur Messung dieser Variablen möglichst auch die von ihnen eingesetzten Tests

zu übernehmen. Dementsprechend werden hier also die Variablen konkreter durch das umschrieben, was die Testautoren mit ihren operationalen Definitionen darunter verstehen. Dazu sei auf deren Arbeiten verwiesen bzw. auf die Beschreibung der Messverfahren im Methodikteil dieser Arbeit.

Soviel zur Begriffsverwendung bzw. Definition.

In BORCHERT & MASENDORFs Untersuchung wurde die Schulart nicht kontrolliert. Ihre Vermittlungshypothese soll offenbar unabhängig von der "Leistungsgruppen-Zugehörigkeit" innerhalb des segregierten Schulsystems, also unabhängig von der von den Schülern besuchten Schulart zutreffen. Der Schüler im segregierten Schulsystem kann sich aber durchaus als Mitglied einer Schulart verstehen und das segregierte Schulsystem zum Bezugspunkt der Einschätzung der eigenen Leistungsfähigkeit machen. Bisher hatten wir die postulierten Kausalzusammenhänge zwar nur auf die Förderstufe bezogen, doch vermuten wir, dass sie ganz unabhängig davon gelten, ob man (d.h. die Schüler) die gruppenbezogenen Leistungsvergleiche innerhalb des traditionellen segregierten oder innerhalb des integrierten Schulsystems anstellt. Solange nämlich die schulische Leistungsfähigkeit primär anhand von Sprach- und Rechenleistungen eingeschätzt wird, scheinen sich auch Sonder-/Haupt-Schüler innerhalb eines segregierten Schulsystems verglichen mit Realschülern bzw. Gymnasiasten sowie C-Kurs-Schüler innerhalb eines integrierten Schulsystems verglichen mit B- bzw. A-Kurs-Schülern ihres Schuljahrganges in einer ähnlich schwierigen Lage zu befinden.

Dass hier Gemeinsamkeiten bestehen, erscheint auch von der Theorie der Bezugsgruppenprozesse her plausibel. Wenn für Schüler unterer Leistungsgruppen deutlich sichtbar

- nur wenige Schüler ihrer Umgebung in ihre untere (sozial niedrig bewertete) Leistungsgruppe geraten,
- die Zugehörigkeit zur unteren Leistungsgruppe, als schulischer Misserfolg relativ klar und öffentlich vergleichend dargestellt, zu einer Dauereinrichtung wird,

dann, so vermuten wir, wird die niedrige Leistungseinstufung vom Schüler nicht mehr sachbezogen im Sinne der beabsichtigten Förderung, sondern selbstwertbezogen im Sinne eines sozialen Vergleichs interpretiert: Es entsteht das Gefühl, zu den als nicht sehr leistungsfähig geltenden Schülern zu gehören.

Allerdings nehmen wir an, dass der vermutete Struktureinfluss sich stärker noch in einer quasi-streaming organisierten Förderstufe als im segregierten Schulsystem (mit den "Schularten" als den zu vergleichenden Leistungsgruppen) bemerkbar machen müsste.

Wahrscheinlich deutlicher als im segregierten Schulsystem wird den Schülern unterer Leistungsgruppen in der Förderstufe zu verstehen gegeben, das Aufsteigen in höhere Leistungsgruppen sei erwünscht und werde auch gefördert. Leichtere Ueberschaubarkeit und stärkerer Oeffentlichkeitscharakter der Kurseinstufungen könnte daher dem Schüler eher zeigen, was er "wert" ist. Zumal der unteren Leistungsgruppe wird im Unterschied zu derjenigen im segregierten Schulsystem (Sonder-/Hauptschule) ja ihre faktisch ungünstige Stellung in der schulischen Statushierarchie durch unterrichtsorganisatorischen und auch räumlichen Wechsel tagtäglich deutlich vor Augen geführt.

Im quasi-streaming-Kurssystem könnte dem Schüler der unteren Leistungsgruppe im Laufe der Zeit noch unmittelbarer bewusst werden, dass für ihn nur geringe Aufstiegschancen bestehen. Mag er anfangs die Kurseinstufung noch als Möglichkeit zur Förderung im Sinne eines baldigen (Wieder-)Aufstiegs und nicht als endgültige Beurteilung seiner Leistungsfähigkeit verstehen, so könnte jedoch im Laufe gleichbleibend niedriger Einstufungen ein allmählicher Wandel in seiner Situationsdefinition eintreten: Am Ende einer relativ stabil quasi-streaming organisierten Förderstufe, so vermuten wir, werden die höheren Kursniveaus für unerreichbar gehalten, d.h. der C-Kurs-Schüler glaubt nicht mehr daran, durch eigene Anstrengungen und eigenen Ein-

satz die schulische Leistungsposition verbessern zu können. Er erfasst, dass die soziale Umwelt die Unabänderlichkeit seiner Kurseinstufung nicht mehr auf zeitvariable Ursachenfaktoren zurückführen kann, sondern im Gegenteil zu der Ansicht kommen muss, seine schulische Leistungsfähigkeit sei gering einzuschätzen.

Fassen wir die bisherigen Ueberlegungen in provisorische, d.h. in noch nicht prüfbar formulierte, Hypothesen zusammen:

1. Wir vermuten, dass am Ende einer streaming-ähnlich organisierten Förderstufe die Zusammenhänge zwischen der sozialen Herkunft des Schülers und der Lehrerunterstützung sowie zwischen sozialer Herkunft und schulischer Anstrengungsbereitschaft bei Kontrolle der schulischen Leistungsfähigkeit bzw. der Kurszugehörigkeiten in Mathematik und Sprachen verschwinden.

2. Darüber hinausgehend vermuten wir, dass bei Kontrolle der schulischen Leistungsfähigkeit bzw. der Kurszugehörigkeiten auch die Zusammenhänge zwischen der Testintelligenz und der Lehrerunterstützung sowie zwischen der Testintelligenz und der Anstrengungsbereitschaft verschwinden. Wir erwarten also, dass die Einflüsse der sozialen Herkunft des Schülers und seiner Testintelligenz auf seine Anstrengungsbereitschaft entscheidend durch seine Kurszugehörigkeiten vermittelt werden.

Die empirische Ueberprüfung dieser Hypothesen stellt einen Test der Vermittlungshypothese von BORCHERT & MASENDORF dar, liefert zugleich aber auch die Möglichkeit zum Test dessen, was die Förderstufe im Hinblick auf die von ihr selbst gesetzte Zielvorstellung "Fördern statt Auslesen", bezogen auf die Anstrengungsbereitschaft, leistet.

2. METHODIK

2.1. Modellspezifikation

2.1.1. Vorbemerkungen

Zur Konzeption einer Kausalhypothese gehört immer auch die explizite Angabe der vereinfachenden Bedingungen, unter denen sie gilt, d.h. die Postulate der in Frage stehenden Theorie müssen ein "abgeschlossenes" System bilden, damit die zu überprüfenden Aussagen überhaupt deduziert werden können. Dabei macht es prinzipiell keinen Unterschied, ob die Schliessung des Systems auf experimentellem oder auf analytischem Wege plausibel gemacht wird. Wichtig ist nur, dass eine ausreichende Kontrollmöglichkeit gegeben ist.

Da wir nicht induktiv verallgemeinern, sondern nur vorgegebene allgemeine Hypothesen testen wollen, kommt es uns vorrangig darauf an, die interne Validität (sensu CAMPBELL & STANLEY, 1963) der Untersuchung zu sichern. Einen einzig richtigen Weg dazu gibt es hier nicht. Denn die unabhängigen Variablen dieser Untersuchung sind zwar veränderbar, doch da in den "natürlichen" Ablauf des elterlichen und schulischen Erziehungsgeschehens nicht planmässig und absichtlich eingegriffen werden darf, kann experimentelle Kontrolle nur bis zu einem den vorfindlichen Gegebenheiten angemessenen, d.h. die bestehenden Sozialbezüge nicht störenden Ausmasse ausgeübt werden.

Wenngleich in dieser Untersuchungssituation eine Randomisierung der Versuchspersonen nicht möglich ist, so gehen wir doch davon aus, dass das Problem der internen Validität hier verhältnismässig gut gelöst werden kann.

Zunächst ist die zu untersuchende Population zeitlich und lokal so zu begrenzen, dass eine Reihe von Alternativerklärungen bereits vorab als weniger wahrscheinlich ausgeschaltet wird. Sodann lassen sich auf statistischem Wege Alternativerklärungen prüfen, die wir wegen der fehlenden Manipulations- bzw. Randomisierungsmöglichkeit erst a posteriori, je nach dem Ausgang der Untersuchung, als wenig plausibel ausschliessen können.

Im Einzelnen dürfte eine für unsere Fragestellung ausreichende Kontrolle in der Kombination folgender Verfahren liegen:

1. Kontextbezüge oder Rahmenbedingungen des Kurssystems als Ganzes sollten für alle untersuchten Schüler weitgehend die gleichen sein.
2. Kontrolle von Selektionseffekten bei der Auswahl der Schüler.
3. Statistische Kontrolle von als "störend" vermuteten Personenmerkmalen.
4. Kontrolle von Messinstrumenten-Fehlern.
5. Kontrolle der Repräsentativität von Messungen.

Die Punkte 3-5 beziehen sich auf a priori als bedeutsam vermutete, aber experimentell nicht kontrollierbare Störvariable.

In diesem Zusammenhang sollte man vielleicht besonders betonen, dass wir an sämtlichen notwendigen Bedingungen der Anstrengungsbereitschaft, die erst in ihrer Gesamtheit die im strengen Sinn hinreichende Bedingung der Anstrengungsbereitschaft darstellen, nicht interessiert sind. Es geht in dieser Arbeit also nicht darum, einen möglichst hohen Varianzanteil der Anstrengungsbereitschaft aufzuklären. Kausalmodelle oder Erklärungen sind immer unvollständig insofern, als stets trivialerweise gesagt werden kann, es gebe Variable, die sie nicht berücksichtigen.

2.1.2. Auswahl der Versuchspersonen

Die interne Validität der Untersuchung kann durch die Unterschiedlichkeit der Förderstufen(-Schulen) in Frage gestellt werden. Nicht an allen Förderstufen bestehen z.B. die gleichen Unterweisungsmethoden, Lehrmittelausstattungen, usw. Die Wirkung dieser Variablen kann kontrolliert werden, wenn wir Schüler desselben Schuljahrganges in nur einer Förderstufe untersuchen. Damit wird eine gewisse Isolation der uns interessierenden Variablen im Felde erreicht.

Es ist klar, dass man zur Widerlegung einer als allgemeingültig gedachten Hypothese keine repräsentativen Stichproben von Förderstufen braucht bzw. nicht auf grosse Versuchspersonen-Zahlen angewiesen ist. Wenn man an nur einer Förderstufe zeigen kann, dass eine konkurrierende Hypothese, die die erwähnte schulische Strukturvariable wesentlich enthält, plausibler ist, kann man die so allgemein formulierte Hypothese als vorläufig widerlegt betrachten. Doch auch bei einer solchen Eingrenzung der Untersuchungspopulation ist die Auswahl der Versuchspersonen möglichst so zu organisieren, dass die Hypothesen mit zufallskritischen Verfahren überprüft werden können.

Grund- bzw. Auswahlgesamtheit der Untersuchung ist also ein Schuljahrgang am Ende seiner Förderstufenzeit. Aus diesem Schuljahrgang muss eine Zufallsauswahl der Schüler getroffen werden, um die Ergebnisse nicht durch den Störfaktor "systematische Ausfälle" zu gefährden ("Zufallsauswahl" ist hier nicht zu verwechseln mit der "Zufallsaufteilung" der Schüler auf die Kursniveaus, was in unserem Falle ja nicht möglich ist). Vor allem die Verzerrungsmöglichkeit durch Selbstselektion soll hier kontrolliert werden. Dies wird vermutlich dann am besten erreicht, wenn

alle für die Untersuchung in Frage kommenden Schüler, die während eines normalen Schulalltages in der ausgewählten Schule sind, gleichzeitig befragt werden. Es ist dann zu erwarten, dass am Befragungstermin fehlende Schüler aus Gründen nicht befragt werden konnten, die von den Untersuchungsvariablen unabhängig sind ("zufällige" Ausfälle, z.B. Krankheit).

Da wir uns auf den Schuljahrgang nur einer Förderstufe beschränken, müssen wir, wenn die Zufallsauswahl sinnvoll sein soll, auch sicherstellen, dass nicht ein Jahrgang ausgewählt wird, in dem Schüler mit Merkmalen über- bzw. unterrepräsentiert sind, welche in systematischer Beziehung zu mehreren Variablen unserer Untersuchung stehen. Wir gehen davon aus, dass dies einigermassen sichergestellt ist, wenn

1. die ausgewählte Förderstufe eine Pflichtschule bzw. keine Angebotsschule ist. Bei Förderstufen, die sich isoliert im Umkreis traditioneller Schulen befinden (als Angebotsschulen), sind eher Selektionseffekte zu erwarten (z.B. derart, dass dort Schüler mit Lernschwierigkeiten überrepräsentiert sind),
2. die angezielte Schülerpopulation in Bezug auf die soziale Herkunft ungefähr den Grössenverhältnissen in der BRD entspricht. Der Einzugsbereich der ausgewählten "Pflicht-Förderstufe" sollte ökologisch nicht einseitig strukturiert sein (das wäre z.B. der Fall, wenn vorwiegend Kinder von Industriearbeitern die Förderstufe besuchten).

Weitere, auf diese Weise (Punkte 1 und 2, s. S. 30) nicht auszuschaltende, jedoch theoretisch plausible Alternativeinflüsse, die in Personenmerkmalen liegen, können kontrolliert werden, indem sie miterhoben und bei der Hypothesenprüfung, damit sie nicht "stören", rechentechnisch auspartialisiert werden.

2.1.3. Kontrollvariable

Die zur Ueberprüfung des Kausalmodells benötigte geschlossene Darstellung des Variablensystems erscheint gerechtfertigt, wenn die postulierte Variablenstruktur gegenüber zusätzlich eingeführten antezedierenden Variablen invariant bleibt und das Modell gleichzeitig eine rekursive Kausalstruktur hat. Dies gilt um so mehr, als wir keine Feininterpretation der Modellparameter beabsichtigen, wir also nicht wissen wollen, wie stark die kausalen Beziehungen im einzelnen sind (also z.B. zum Zwecke der Vorhersage von Kriteriumsvariablen), sondern nur testen wollen, ob bestimmte direkte Einflüsse bestehen oder nicht (vgl. SIMON & ANDO, 1969).

Die Variable "Geschlecht" wird in der Regel als zu kontrollierende Variable angesehen. Auch erscheint es plausibel, analog zum Lehrerverhalten das Unterstützungsverhalten der Eltern zu kontrollieren, dies vor allem deshalb, weil die Förderstufe bei der Verfolgung ihrer Zielvorstellungen für sich selbst die Entscheidung getroffen hat, die Einstufungen in die Kurse nicht von den jeweiligen Förderungsbemühungen der Eltern abhängig zu machen. Wegen der pädagogischen Veränderungsabsicht der Förderstufe ist damit unter der Voraussetzung eines relativ stabilen quasi-streaming organisierten Kurssystems per Entscheidung die Kausalrichtung "Kurszugehörigkeit ⟶ Elternunterstützung" impliziert.

2.1.3.1. Unterstützungsverhalten der Eltern

Der Einfluss der sozialen Herkunft auf die Anstrengungsbereitschaft kann möglicherweise über das Unterstützungsverhalten der Eltern vermittelt werden, ohne dass schulische Variablen wie Kurszugehörigkeit oder Lehrerunterstützung eine wesent-

liche Rolle dabei spielen. Mit anderen Worten, die Einführung der Variablen "Elternunterstützung" könnte unsere eigene Hypothese falsifizieren. Damit verschärfen wir die Hypothese von der entscheidenden vermittelnden Variablen "Kurszugehörigkeit": Die Kurszugehörigkeit ist dann die entscheidende vermittelnde Variable, wenn bei ihrer Kontrolle nicht nur die Zusammenhänge zwischen sozialer Herkunft bzw. Testintelligenz einerseits und Lehrerunterstützung bzw. Anstrengungsbereitschaft andererseits verschwinden, sondern auch die Zusammenhänge zwischen sozialer Herkunft bzw. Testintelligenz einerseits und Elternunterstützung andererseits.

Auch bei Berücksichtigung der neuen Variablen erwarten wir, dass die Anstrengungsbereitschaft nur von den schulischen Variablen beeinflusst wird. Die Hypothesenverschärfung lässt sich in einem Pfaddiagramm so darstellen:

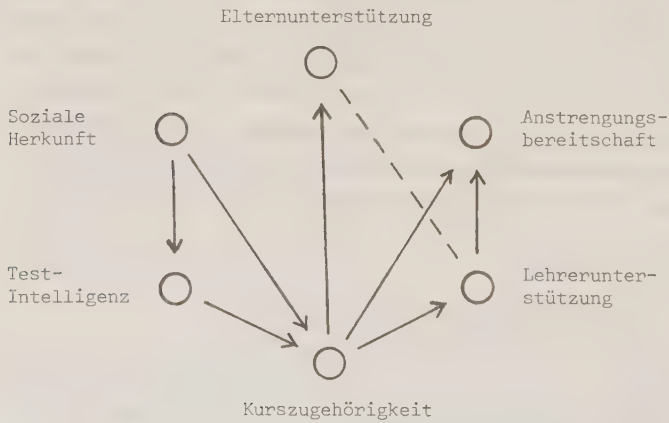

Abb. 6: Gegenüber Abb. 5 um die Kontrollvariable "Elternunterstützung" erweiterte Kausalstruktur zur Erklärung der Anstrengungsbereitschaft: die Kurszugehörigkeit als entscheidende vermittelnde Variable. -------: bislang noch nicht festgelegte Kausalbeziehung

BORCHERT & MASENDORF (1975) haben in ihrer Untersuchung ebenfalls die elterliche Unterstützung, allerdings nur die der Mutter erfasst. Unsere Hypothese steht damit auch zu der spezielleren Vermittlungshypothese BORCHERT & MASENDORFs in Widerspruch, der mütterliche Erziehungsstil und der Erziehungsstil der Lehrer würden den Einfluss der sozialen Herkunft auf die Anstrengungsbereitschaft vermitteln (siehe auch S. 7-9 in dieser Arbeit).

2.1.3.2. Geschlecht

Aufgrund der zahlreichen empirischen Belege, nach denen hinsichtlich der Anstrengungsbereitschaft, auch hinsichtlich der Selbsteinschätzung und Selbsterwartung von Schulleistungen, Geschlechterunterschiede zu bestehen scheinen (z.B. HECKHAUSEN, 1972; VOLLMER & ALMAS, 1974), ist nicht auszuschliessen, dass die Einbeziehung der Variablen "Geschlecht" als exogene Variable unter Umständen die betrachtete Kausalstruktur erheblich verändert.

Bisher wurde die Frage der Kontrolle so angegangen, dass mehr zugunsten interner Validität der Untersuchung gearbeitet wurde: durch Homogenisierung des Untersuchungsumfeldes, Zufallsauswahl der Versuchspersonen und statistische Kontrolle, letzteres in dem Sinne, dass vermutete Störvariable in das Kausalmodell "hereingeholt", also wie explizite Modellvariable behandelt werden. Um nun der Generalisierbarkeit unserer Ergebnisse wieder etwas mehr Raum zu verschaffen, wird an dieser Stelle eine andere Form statistischer Kontrolle vorgeschlagen.

Die beste Sicherung gegen Zufallsergebnisse bietet die Wiederholung der Analyse an weiteren Stichproben aus derselben Population. Eine Variante dieses Vorgehens liegt darin, eine vorliegende Stichprobe zufällig in zwei Teilstichproben aufzuteilen und zu prüfen, ob die in der einen Teilstichprobe gewonnenen Ergebnisse in der anderen Teilstichprobe reproduzierbar sind. Im Sinne der Falsifikationstheorie POPPERs (1969[3]) empfehlens-

werter wäre es, die Stichprobe nicht zufällig, sondern bewusst nach einer Variablen aufzuteilen, von der ein Einfluss auf die Untersuchungsergebnisse erwartet wird. Die Variable, der wir diesen Einfluss am ehesten zutrauen, ist das Geschlecht der Versuchspersonen.

Wir postulierten unter anderem eine Abhängigkeit der Elternunterstützung von den Schulleistungen ihrer Kinder (bei vorgegebener Kausalrichtung). Untersuchungsergebnissen von HERRENKOHL (1972) zufolge könnte diese Hypothese höchstens für die Gruppe der Jungen gelten. Auf die Mädchen liesse sich dies nicht übertragen: In den Schulleistungen gute wie schlechte Mädchen berichten generell von geringerer elterlicher Anteilnahme und Unterstützung. Bei den Jungen dagegen fühlen sich diejenigen mit den guten Schulleistungen stärker als die mit den schlechten Leistungen von den Eltern in ihrem Leistungsstreben unterstützt.

Dieselbe Geltungseinschränkung der Hypothese kann sich ergeben, wenn man die Person des Lehrers in die Ueberlegungen einbezieht. Einige Autoren berichten, dass das Arbeits- und Sozialverhalten der Mädchen in der Schule den Bemühungen der Lehrer eher entgegenkomme als das der Jungen: Im Gegensatz zu vergleichbaren Jungen zeigen auch die Mädchen mit den schlechteren Schulleistungen ein günstigeres Arbeitsverhalten (ANDERSON, 1970). Lehrer neigen daher auch dazu, leistungsschwache Mädchen positiver zu beurteilen als leistungsschwache Jungen (BROPHY & GOOD, 1970). Mädchen erhalten generell vom Lehrer mehr Lob als Jungen (OTTO, 1970).

Diese Befunde mögen zum vorläufigen Beleg von Geschlechtsunterschieden in diesem speziellen Sektor genügen. Weitere Resultate anzuführen wäre auch wenig sinnvoll, zumal es oft schwierig und auch problematisch ist, Einzelergebnisse verschiedener Studien im nachhinein innerhalb einer umfassenderen Theorie kausal zu interpretieren.

Zumindest lassen die wenigen hier berichteten Geschlechtsunterschiede es als angezeigt erscheinen, den auf sämtliche hier betrachteten Variablenbeziehungen noch nicht geklärten Geschlechtseinfluss zu kontrollieren.

Anstatt nun das bisher postulierte Kausalmodell um eine exogene Variable "Geschlecht" zu erweitern, werden wir es in jeder Geschlechtergruppe separat überprüfen. Praktisch läuft dies auf eine Kreuzvalidierung der Ergebnisse, auf eine Ueberprüfung der externen Validität im Sinne von CAMPBELL & STANLEY (1963) hinaus. Wenn sich die Untersuchungshypothese in beiden Geschlechtergruppen gleichermassen bewährt, ist die Möglichkeit der Generalisierbarkeit des Ergebnisses erheblich gestiegen.

2.1.4. Erfassung der Variablen

1. Schulische Anstrengungsbereitschaft: Zur Erfassung der Anstrengungsbereitschaft verwenden wir den "Anstrengungsvermeidungstest" (AVT) von ROLLETT & BARTRAM.

"Anstrengungsvermeidung" betrachten wir hier als das Gegenteil von "Anstrengungsbereitschaft". ROLLETT & BARTRAM legten den AVT bereits 1972 vor, der damals noch aus 35 Items bestand (Ich danke den Testautoren für ihr Einverständnis zur Anwendung ihres Fragebogens). Später wurde dieser Test auf 20 Items (siehe im Anhang) reduziert, die gemäss dem logistischen Modell von RASCH skaliert sind. Die RASCH-homogenen Items werden mit den Antwortmöglichkeiten "stimmt" und "stimmt nicht" vorgegeben.

Der Test soll die Neigung erfassen, schulischen Leistungsanforderungen routinemässig aus dem Wege zu gehen. Die Autoren baten mehrere Lehrer, eine Sammlung jener Ausreden anzulegen, die 12 - 14jährige Schüler dann vorbringen, wenn sie sich vor Leistungsanforderungen zu "drücken" versuchen. ROLLETT & BARTRAM zufolge unterscheidet sich der Anstrengungsvermeider vom Misserfolgsängstlichen dadurch,

> "...dass er auch dann keinen Ehrgeiz in der ihn nicht interessierenden Leistungssituation entwickelt, wenn er von seinen Fähigkeiten her gesehen dazu in der Lage wäre" (1977, S. 5)

Den Anstrengungsvermeider kennzeichnen Interesselosigkeit, Resignation, manchmal auch oppositionelle Verweigerung (z.B. Ablehnung, Agression) gegenüber Leistungsanforderungen. Schulischer Erfolg erscheint ihm eher gleichgültig.

Die von den Testautoren geschätzte interne Konsistenz beträgt $r_{tt} = 0,80$. Untersuchungsergebnisse an einer Hauptschüler-Stichprobe (7. und 8. Schuljahrgang) stützen die Vermutung, dass der Anstrengungsvermeidungstest einen wesentlichen Beitrag zur Erklärung des Lernerfolgs zu leisten vermag (ROLLETT & BARTRAM, 1973). Ferner korreliert der Test signifikant mit den Schulnoten in Rechnen ($r = 0,25$) und Deutsch ($r = 0,31$). Mit den Marburger Unterstützungs-Skalen korreliert er (nicht signifikant) zu $r = 0,15$ (Vater-Version) bzw. zu $r = 0,17$ (Mutter-Version).

In der vorliegenden Untersuchung wird die Auswertungsrichtung der 20 RASCH-skalierten Items umgekehrt: Hier ergibt die Summe der "stimmt-nicht" - Antworten den Skalenwert. Dieser reicht von 0 - 20, wobei höhere Skalenwerte grössere Anstrengungsbereitschaft bedeuten.

2. Soziale Herkunft: Die soziale Herkunft wird erfasst mit Hilfe von BAUERs (1972) "Verfahren zur Messung des für das Bildungsverhalten relevanten sozialen Status" (BRSS). Aus Gründen der Anonymitätsabsicherung konnten Informationen zur sozialen Herkunft eines Schülers von uns selbst nicht eingeholt werden. Diese wurden uns vom "Deutschen Institut für Internationale Pädagogische Forschung" (DIPF) in Frankfurt/M. ohne Namennennung in Form einer einzigen Zahl für jeden Schüler (sog. "Kennziffer") übermittelt.

BAUERs Verfahren stellt eine Modifikation des "Kölner Index" von SCHEUCH (1961) dar. Wichtigstes Merkmal dieser Modifikation ist, dass der Schulbildung (von Vater und Mutter zusammengenommen) gegenüber der beruflichen Position des Vaters das grössere Gewicht bei der Index-Bildung eingeräumt wird. Daher erscheint in unserem Untersuchungszusammenhang auch die Verwendung des Attributs "sozio-kulturell" anstelle von "sozial" in Verbindung mit dem Begriff "Herkunft" angebracht.

Der Index von BAUER setzt sich zusammen aus den Komponenten "Beruf des Vaters", "Schulbildung des Vaters" und "Schulbildung der Mutter". Die diesen Komponenten jeweils zugeordneten Zahlenwerte werden additiv zu einem Indexwert kombiniert. Der gesamte Indexwerte-Bereich wird in sechs Abschnitte bzw. soziale Schichten unterteilt. Die Namen dieser sozialen Schichten, mit den zugehörigen Kennziffern in Klammern, lauten: Oberschicht (1) - Obere Mittelschicht (2) - Mittlere Mittelschicht (3) - Untere Mittelschicht (4)- Obere Unterschicht (5) - Untere Unterschicht (6).

Zur Reliabilität seines Verfahrens macht BAUER in seinem Manual von 1972 keine numerischen Angaben. Mit anderen unabhängigen Indikatoren der sozialen Herkunft weist der

BRSS folgende Korrelationen auf: Mit der SSE-Skala (Soziale Selbsteinschätzung) von KLEINING & MOORE (1960) von r = 0,81, mit einem von PEISERT & DAHRENDORF (1967) vorgeschlagenen Verfahren zur Messung des Sozialstatus von r = 0,82. Aus der Höhe dieser Korrelationen kann man ableiten, dass der BRSS eine hohe interne Validität (im Sinne der Korrelation des BRSS mit dem gemeinsamen Faktor) besitzt. Allerdings liegen noch zu wenige Angaben darüber vor, ob man dem BRSS auch befriedigende externe Validität im Sinne der Uebereinstimmung mit entsprechendem Bildungsverhalten (konkurrente und/oder Vorhersage-Validität) attestieren kann. Man kann aber davon ausgehen, dass die Kurszugehörigkeit (bzw. der objektive schulische Status) eines Schülers enger mit dem BRSS korreliert als mit einem Index der sozialen Herkunft, der die sozio-ökonomische Komponente favorisiert. Dies erscheint plausibel, wenn man sich vor Augen hält, dass ein hoher Bildungsgrad der Eltern oft auch bei weniger gutem beruflichen Status und/oder finanzieller Lage der Familie durchschlägt, den Kindern zu einer guten Schulausbildung (bzw. Kurseinstufung) zu verhelfen.

Für die Auswertung dieser Untersuchung wird die Schlüsselrichtung der vom DIPF/Frankfurt mitgeteilten Kennziffern umgekehrt, d.h. der unteren Unterschicht wird die "1" zugeordnet, der oberen Unterschicht die "2", usw., bis zur Oberschicht, die die "6" erhält.

3. Allgemeine intellektuelle Leistungsfähigkeit: Die allgemeine intellektuelle Leistungsfähigkeit erfassen wir mit dem "Grundintelligenztest-Skala 2" (CFT2) von WEISS (1972). Dieser Test ist eine deutsche Version des "Culture Fair Intelligence Test - Scale 2" von R. B. CATTELL. Er soll einen sprachfreien Generalfaktor der Intelligenz messen.

Der CFT2 in der deutschen Fassung besteht aus zwei Testteilen, die den gleichen Aufbau haben, u.a. dieselben vier Untertests: Series, Classification, Matrices und Topology. Diese vier Untertests korrelieren mit dem Generalfaktor zwischen r = 0,78 und 0,83.

Die interne Konsistenz des gesamten Tests beträgt r_{tt} = 0,93. Der CFT2 korreliert mit schulischen Leistungen zwischen r = 0,36 bis 0,52, für die Note in Deutsch etwas geringer als für die in Mathematik. Mit sprachabhängigen Wortschatz- und Rechtschreibtests korreliert er wesentlich niedriger (r = 0,09) als mit weniger sprachgebundenen Testverfahren wie z.B. dem Matrizentest von RAVEN (r = 0,58) oder dem Frankfurter Analogie-Test (r = 0,68). Allerdings können wir der von WEISS daraus abgeleiteten Ansicht nicht ohne weiteres beistimmen, mit dem CFT2 könne das "Intelligenzpotential" eines Schülers unbeeinflusst von seiner soziokulturellen Herkunft erfasst werden (1972, S. 3 bzw. S. 9). Den Ausführungen von MERZ & STELZL (1977, S. 82-83) zufolge muss man im Gegenteil eher damit rechnen, dass sich das soziokulturelle Milieu in Leistungen zum vermeintlich "kulturfreien" oder "kulturfairen" Test stärker niederschlägt als in den Leistungen zu den sprachgebundenen oder verbalen Tests. Ein vorwiegend aus soziokulturellen Indikatoren gebildeter Index der sozialen Herkunft wie z.B. der hier verwendete BRSS von BAUER sollte daher mit dem CFT2 eine nicht unbeträchtliche gemeinsame Varianz aufweisen.

Auch können wir die von WEISS (1972, S. 18) vertretene Auffassung nicht akzeptieren, der CFT2 messe die "allgemeine intellektuelle Leistungsfähigkeit". Wir begegnen dem gleichen Problem wie bei der Definition der schulischen Leistungsfähigkeit (s. S. 14-23 in dieser Arbeit): Auch hier ist es nicht möglich, die "intellektuelle Leistungsfähigkeit"

ohne Berücksichtigung gesellschaftlich abhängiger Werthierarchien zu definieren. Mit HUSÉN (deutsch 1977, S. 55) sind wir der Meinung, dass das mehr oder weniger explizite Ziel der meisten konventionellen Intelligenztests die Messung der "Fähigkeit zur Durchführung abstrakter Operationen mit Hilfe von Symbolen" (übersetzt durch PFEFFER) ist. "Intelligenz anderer Art", z.B. die Fähigkeit zum Umgang mit konkreten, greifbaren Dingen, also handwerkliches Geschick, kann bei diesen Tests in nur sehr geringem Masse zu Buche schlagen.

Wenn wir an bestimmten Stellen dieser Arbeit die Intelligenzleistung im CFT2 als Indikator der "(allgemeinen) intellektuellen Leistungsfähigkeit" bezeichnen, so folgen wir testpsychologischem Sprachgebrauch. Stets ist bei dieser Redeweise jedoch die Einschränkung mitzubedenken, dass wir mit dem CFT2 eigentlich mehr die Fähigkeit zu abstraktem Denken erfassen, eine Fähigkeit, die sich sprachlicher oder nicht-sprachlicher Mittel bedienen kann. Um diesen Bedenken auch in der weiteren Darstellung einen gewissen Ausdruck zu geben, wollen wir, soweit angebracht, anstelle von "intellektueller Leistungsfähigkeit" auch von "Testintelligenz" sprechen.

Aus Zeitgründen wird nur der erste Testteil des CFT2 angewendet. Für Gruppenvergleiche reicht das aus, zumal die von WEISS geschätzte interne Konsistenz des so gekürzten Tests noch 0,86 beträgt. Allerdings wurde dies nur für die Gruppe 11jähriger Schüler aus dem 5. Schuljahrgang festgestellt. Noch nicht geklärt ist, ob dieser Wert auch für die Gruppe der 12-14jährigen Schüler gilt, mit denen wir es in unserer Untersuchung hauptsächlich zu tun haben.

Die Rohpunktwerte des ersten Testteils laufen von minimal 0 bis maximal 41. Die Auswertungsrichtung ist hier so, dass höhere Rohpunktwerte bessere Intelligenzleistungen bedeuten. Um bei der Testdurchführung ein Abschauen zu verhindern, werden die Parallelformen A und B vorgegeben.

4. Schulische Leistungsfähigkeit: Die schulische Leistungsfähigkeit unserer Versuchspersonen erfassen wir durch ihre Kurszugehörigkeiten in Mathematik und Sprachen, wie sie am Ende der Förderstufenzeit bestehen. Wie bereits in der Einführung begründet, erscheint es bei quasi-streaming organisierten Förderstufen gerechtfertigt, die am Ende der Förderstufenzeit bestehenden Kurszugehörigkeiten als (institutionell definierte) Indikatoren der schulischen Leistungsfähigkeit zu betrachten. Die Berücksichtigung nur dieses Zeitpunktes im Rahmen des vorgegebenen Kausalmodells setzt jedoch voraus, dass die Kurszugehörigkeiten, genauer: das quasi-streaming organisierte Kurssystem über die Förderstufenzeit hinweg relativ stabil war (nach bisher vorliegenden Beobachtungen war dies für die meisten Förderstufen-Schulen zu erwarten).

Auch hier wollen wir eine dem Vorhergehenden analoge Vereinbarung treffen: Um im Laufe unserer Erörterungen nicht der Gefahr eines pars-pro-toto-Fehlschlusses zu unterliegen, wollen wir im folgenden mit dem Begriff "schulische Leistungsfähigkeit" möglichst sparsam umgehen, d.h. wir werden stattdessen öfter den "konkreteren" bzw. "genaueren" Ausdruck "Kurszugehörigkeit(en)" gebrauchen. Doch ist dabei im Auge zu behalten, dass mit der Verwendung des Begriffs "Kurszugehörigkeiten" hier auch das Wissen um seine allgemeinere, institutionelle Bedeutung verknüpft ist.

Den drei möglichen Kurszugehörigkeiten werden in der Auswertung folgende Zahlenwerte zugeordnet:
höchstes Kursniveau: 3, mittleres Kursniveau: 2, und niedrigstes Kursniveau: 1.

5. Eltern-Unterstützung: Die Eltern-Unterstützung erfassen wir mit Hilfe der Marburger Erziehungsstil-Skalen von HERRMANN et al. (1968) und STAPF et al. (1972). Vater- und Mutterversion der Unterstützungs-Skalen bestehen aus jeweils 15 Items (siehe im Anhang), welche in den Formulierungen teilweise übereinstimmen. Zu jedem Item stehen fünf Antwortmöglichkeiten zur Verfügung. Der Skalenwert, für die Vater- und Mutterversion getrennt erhoben, ergibt sich aus der Summe der von den Autoren den Itemantworten zugeordneten Gewichtszahlen.

Beide Skalen sind eindimensional, im faktorenanalytischen Sinne wie auch im Sinne des RASCH-Modells, und sie korrelieren hoch positiv miteinander ($r = 0,61$). Aehnlich hoch fällt die Korrelation in einer Untersuchung von LUKESCH & TISCHLER (1975) aus ($r = 0,57$). Für Jungen berichten sie eine Korrelation von 0,60, für Mädchen eine von 0,54.

Die split-half-Zuverlässigkeit der Vater-Unterstützungsskala liegt bei 0,87, die der Mutter-Unterstützungsskala bei 0,80 (STAPF et al., 1972, S. 104).

Zu beachten ist, dass hier nicht etwa die Eltern selbst über ihr Erziehungsverhalten Auskunft geben, sondern dass hier Kinder berichten, wie sie das Erziehungsverhalten der Eltern ihnen gegenüber erleben. Zur Erfassung der elterlichen Unterstützung bietet sich auch die Möglichkeit an, die Daten von den Eltern direkt zu erheben. Der Erlebnisperspektive der Kinder wird hier jedoch der Vorrang gegeben, weil wir uns auf den Standpunkt stellen, dass es bei der Erklärung der Anstrengungsbereitschaft vorwiegend - andere Autoren (z.B. AUSUBEL, 1958, S. 356) würden hier sagen: "letztlich nur" - darauf ankommt, wie das Kind das Ermunterungs- und Unterstützungsverhalten der Eltern erlebt. Die Brauchbarkeit dieses Vorgehens belegen die von STAPF et al. (1972) durchgeführten Validitätsstudien. Einige Ergebnisse daraus, die mit

unserer Fragestellung in unmittelbarerem Zusammenhang stehen: eher Kinder mit guten Leistungen in der Schule (Rechnen, Heimatkunde) und im Intelligenztest (HAWIK-Subtest: "Allgemeines Verständnis") erleben ihre Eltern als unterstützend, ebenso Kinder aus Familien mit einem höheren sozioökonomischen Status.

In beiden Skalenversionen bedeuten höhere Skalenwerte (Extremwerte: 15 und 75) stärkere Unterstützung.

6. Lehrer-Unterstützung: Hier soll Analoges gelten, wie es bezüglich der Erfassung der Eltern-Unterstützung gesagt wurde. Wir gehen davon aus, dass nur die Schüler Informationen darüber geben können, wie das Erziehungsverhalten der Lehrer bei ihnen "ankommt". Die Auswirkungen des eigenen Verhaltens auf die Schüler können Lehrer oft nicht richtig einschätzen (z.B. TAUSCH & TAUSCH, 1968, S. 96, S. 117). Die Lehrer-Unterstützung wollen wir daher mit einer von HEINRICH (1974) vorgelegten Skala erfassen.

Diese Skala, bestehend aus 15 Items (siehe im Anhang) wurde zusammen mit einer Lehrer-Strenge-Skala in Anlehnung an die Marburger Erziehungsstil-Skalen entwickelt. Die Konstruktion dieser Skalen entspricht dem von HERRMANN et al. (1968) vorgegebenen theoretischen Konzept zum elterlichen Erziehungsstil offenbar besser als andere Varianten der Marburger Skalen im schulischen Bereich (z.B. Skalen von MASENDORF & TSCHERNER, 1973; Skalen von BORCHERT, 1974): Bei einer von HEINRICH (1974) durchgeführten Faktorenanalyse seiner Unterstützungs- und Strenge-Items erwiesen sich "Strenge" und "Unterstützung" auch im schulischen Bereich als faktoriell unabhängige Dimensionen des erlebten Erzieherverhaltens. Diese Uebereinstimmung mag vielleicht nicht mehr so zu überraschen, wenn man sich vor Augen hält, dass die meisten der von HEINRICH vorgelegten Items ein formulierungsgleiches bzw. ziemlich formulierungsähnliches Pendant in den Marburger Skalen haben.

Von seiner Lehrer-Unterstützungsskala berichtet HEINRICH eine split-half-Reliabilität von 0,83. Er fand ausserdem das mit unserer Hypothese verträgliche Ergebnis, dass Kinder unterschiedlicher sozialer Herkunft vom Lehrer nicht unterschiedlich stark unterstützt werden (1974, S. 540).

Für unsere Darbietung haben wir die Items geringfügig umformuliert: Anstatt im Singular erscheint die Person des Lehrers im Plural. Die von HEINRICH ermittelten Itemqualitäten verschlechtern sich dadurch vermutlich nicht. Diese Umformulierung erschien angesichts der Tatsache angebracht, dass Schüler täglich von mehreren Lehrern unterrichtet werden. Oekonomie und Standardisierung des Verfahrens sollten sich dadurch verbessern, zumal es nur darauf ankam festzustellen, welches Unterstützungsverhalten der Lehrerschaft vom Schüler als das ihm gegenüber "typische" erlebt wird.

Der Skalenwert ergibt sich aus der Addition der den Itemantworten zugeordneten Gewichtszahlen (Extremwerte: 15 und 75). Wie bei den Marburger Skalen ist die Auswertungsrichtung hier so, dass höhere Skalenwerte stärkere Unterstützung bedeuten.

Die Hereinnahme der Lehrer-Skala von HEINRICH verschärft möglicherweise ein bisher eher latent vorhandenes Problem. Die Mutter- und Vaterversion der Marburger Unterstützungsskalen sind einander formulierungsähnlich. Mit der Lehrer-Skala kommt ein weiteres formulierungsähnliches Verfahren hinzu. Jetzt kann nicht mehr ohne weiteres davon ausgegangen werden, dass die Anwendungsbedingungen der Lehrer/Eltern-Skalen unabhängig voneinander sind. Je formulierungsähnlicher die verwendeten Messverfahren sind, genauer: je deutlicher die vorgegebene Formulierungsähnlichkeit der Messverfahren von den Versuchspersonen wahrgenommen wird, desto eher könnten diese eigene Hypothesen bilden. Auf Messinstrumente, die verschiedene Variablen erfassen sollen, wird unter Umständen gleich rea-

giert, obwohl man sich in den zu erfassenden Variablen unterscheidet. Mit anderen Worten, es ist damit zu rechnen, dass formulierungsähnliche Messverfahren, wenn sie zusammen angewendet werden, eine systematische Antworttendenz (sog. response set) bewirken.

Die angesprochene störende Variable kann hier experimentell nicht konstant gehalten werden. Unter bestimmten Zusatzbedingungen lässt sich dieses Problem aber analytisch lösen. Diese Zusatzbedingungen sind:

1. Ungemischte Darbietung der Verfahren, d.h. der drei Unterstützungsskalen,
2. mindestens zwei Verfahren (Indikatoren) je zu erfassender Variable (d.h. Eltern-Unterstützung, Lehrer-Unterstützung).

Zur ersten Zusatzbedingung: Die Gefahr der angesprochenen Antworttendenz scheint uns eher gegeben, wenn die Items der betroffenen Verfahren gemischt und nicht blockweise getrennt als selbständige Einheiten, eventuell mit grösserem zeitlichen Abstand, dargeboten werden. Zum anderen hätte eine gemischte Darbietung hier einen schwerwiegenden analytischen Nachteil: Es kann nicht kontrolliert werden, ob sich die vermutete Antworttendenz tatsächlich ausgewirkt hat. Der Störeffekt wäre unidentifizierbar in den Variablenbeziehungen enthalten. Bei einer ungemischten Darbietung dagegen wäre die Störung als Sequenzeffekt ("carry-over-effect") fassbar. Einer Antwortverzerrung durch Formulierungsgleichheit bzw. -ähnlichkeit kann nämlich nur das in der Darbietung nachfolgende Messverfahren unterliegen. Im Rahmen einer expliziter ausformulierten Messinstrumenten- oder Messfehler-Theorie kann man diese Kenntnis vorteilhaft zur (analytischen) Kontrolle des Störeffekts nutzen.

Zur zweiten Zusatzbedingung: Die Annahme, dass systematische Instrumenten- oder Mess-Fehler eine Rolle spielen, lässt sich überprüfen, indem man mehrere Indikatoren für dieselbe Variable verwendet und schaut, wie sich dieses Verfahren multipler Indikatoren bewährt. Systematische Messfehler sind in diesem Zusammenhang um so eher auszuschliessen, je häufiger das gleiche Ergebnis mit unterschiedlichen, d.h. hier: nicht-formulierungsähnlichen Messinstrumenten erzielt wird. Da wir die elterliche Unterstützung bereits zweifach messen, bedeutet das, dass zur Erfassung der Lehrer-Unterstützung zumindest ein zweiter Indikator zugezogen werden muss, welcher Itemformulierungen des mitverwendeten Indikators nicht wiederholt.

Diese Anforderung schien uns die im folgenden dargestellte Skala von BRUSTEN & HURRELMANN (1973) zu erfüllen. Die Autoren verwendeten ihre Skala zur Erfassung der - wie sie es ausdrücken - "Einschätzung des Sympathiestatus bei den Lehrern" (S. 91). Uns soll diese Skala dazu dienen, die Lehrer-Unterstützung unabhängig von den Anwendungsproblemen der Lehrer-Skala von HEINRICH zu erfassen.

Die Skala von BRUSTEN & HURRELMANN besteht aus 5 Items (siehe im Anhang). Aus Diskussionen mit Schülern erstellten die Autoren zunächst einen umfangreichen Item-Katalog, aus dem dann nach Vortests die am höchsten miteinander korrelierenden Items ausgewählt wurden. Weitere teststatistische Daten teilen die Autoren nicht mit. Doch kann man annehmen, dass die Items nur einen Faktor gemeinsam haben. Auch wenn dieser Konstruktion kein ausgearbeitetes Mess- oder Skalierungsmodell zugrundeliegt: Bei Eindimensionalität des Messverfahrens kann die Addition der den Itemantworten zugeordneten Zahlenwerte zumindest ordinale Relationen hinsichtlich der zu erfassenden Variablen liefern. Für die Brauchbarkeit dieser "Skala" (besser: "Index") spricht auch die Vielzahl der im Sinne der Konstruktvalidität positiven

Ergebnisse in der Untersuchung von BRUSTEN & HURRELMANN (1973).

In der vorliegenden Untersuchung wurden die Items mit einem fünfstufigen Antwortschema vorgegeben, das aus folgenden Häufigkeitsangaben bestand: fast immer - oft - manchmal - selten - fast nie. Die von BRUSTEN & HURRELMANN übernommenen Items selbst enthielten zum Teil Häufigkeitsangaben (z.B. "Manchmal sind meine Lehrer richtig gemein zu mir"). Um Mehrdeutigkeiten bei der Beantwortung zu vermeiden, wurden die Häufigkeitsangaben aus den ursprünglichen Itemformulierungen entfernt.

Der Indexwert für den einzelnen Schüler ergibt sich aus der Addition der den Antwortmöglichkeiten zugeordneten Zahlenwerte. Die 5 Antwortmöglichkeiten werden als gleichabständig behandelt, wobei diejenige Antwortmöglichkeit die meisten Punkte (hier: 5) erhält, die den höchsten Sympathiestatus ausdrückt. Es ist noch zu klären, ob die einfache oder gewichtete Summe dieser Punkte gebildet werden soll. Soviel steht jetzt schon fest: Höhere Indexwerte bedeuten in der Auswertung stärkere Lehrer-Unterstützung.

2.1.5. Kausale Ordnung der Variablen

Aus Korrelationen lassen sich keine Kausalrichtungen ableiten. Zwei Konsequenzen ergaben sich daraus für diese Untersuchung:

- die Kausalrichtungen wurden a priori, vor Einsicht in die Daten, festgelegt,
- die Kausalrichtungen wurden anhand von Kriterien festgelegt, die ausserhalb der Daten liegen

Die kausale Ordnung der Variablen lässt sich zum Teil schon anhand der verschiedenen Entstehungszeitpunkte der Variablen begründen. Das betrifft vor allem die Beziehungen zwischen den Variablen soziale Herkunft, Testintelligenz und Kurszugehörig-

keit. Wenn wir z.B. die soziale Herkunft mit der Testintelligenz des Schülers verknüpfen, ist die soziale Herkunft der zeitlichen Abfolge nach früher. Wenn wir die Kurszugehörigkeit mit der Testintelligenz in Beziehung setzen, ist letztere Variable zeitlich früher, usw. Vorausgesetzt wird dabei, dass die individuellen Ausprägungen dieser Variablen über die Förderstufenzeit hinweg relativ stabil bleiben.

In Bezug auf die Variablen Kurszugehörigkeit, Elternunterstützung, Lehrerunterstützung und Anstrengungsbereitschaft liegt die kausale Reihung für den Leser möglicherweise nicht so auf der Hand. Doch ist in diesen Fällen auf die Aufgabenstellung der Untersuchung zu verweisen (u.a. Test von BORCHERT u. MASENDORFs Hypothesen, Test von Zielvorstellungen der Förderstufe). Dementsprechend wurde die Kausalrichtung z.B. mit Kurszugehörigkeit "kausal vor" Eltern-, Lehrer-Unterstützung und Anstrengungsbereitschaft festgelegt. Natürlich ist es auch möglich, die umgekehrte Kausalrichtung zu sehen. Dass wir dies nicht tun, hat seinen Grund unter anderem in der Berücksichtigung von Annahmen, die den Zielvorstellungen der Förderstufe zugrundeliegen. Will man testen, ob die Förderstufe ihre Ziele erreicht oder nicht, dann darf die kausale Reihung der Variablen den in eben diese Zielvorstellungen eingehenden Annahmen nicht von vornherein widersprechen.

Nimmt man das von der Förderstufe angestrebte Ziel "Fördern statt Auslesen" ernst, dann sind mit der Erstzuordnung der Schüler in die Kurse auch nach ihrer Anstrengungsbereitschaft zwei grundsätzliche Annahmen impliziert:

- an die in die Förderstufe "hereingelieferten" verschiedenen Anstrengungsbereitschaften bzw. an die dementsprechend vorgenommenen Personenzuordnungen sollen pädagogische Förderungsbemühungen angepasst werden: Bedingungsmodifikation ist möglich,
- durch angepasste, d.h. differenzierte pädagogische Förderungsbemühungen sollen "hereingelieferte" Anstrengungsbereitschaf-

ten verändert werden: Personenmodifikation ist möglich.

Die Absicht der Bedingungs- und Personenveränderung impliziert unter der Voraussetzung, dass die Kurszuordnung relativ stabil sowie relativ undifferenziert, fachübergreifend, also quasi-streaming ist, die Kausalrichtungen: Kurszugehörigkeit kausal vor pädagogischen Förderungsbemühungen, sowie pädagogische Förderungsbemühungen kausal vor Anstrengungsbereitschaft.

Es ist klar, dass diese kausale Anordnung, so allgemein formuliert, nicht nur für die Förderungsbemühungen der Lehrer, sondern gleichermassen auch für die Förderungsbemühungen der Eltern gilt. Die a priori festgelegte Kausalrichtung: Kurszugehörigkeit kausal vor Elternunterstützung steht darüber hinaus in Einklang mit der im Fördergedanken der Förderstufe enthaltenen Vorstellung, die Kurseinstufung solle nicht von den jeweiligen Förderungsbemühungen der Eltern abhängig gemacht werden.

Mit der a priori festgelegten Kausalrichtung von Kurszugehörigkeit zur Elternunterstützung ist nun keineswegs festgelegt, dass die Kurszugehörigkeit die Elternunterstützung tatsächlich beeinflusst, sondern nur, dass die Möglichkeit dazu besteht, mit anderen Worten: bei festgelegter Kausalrichtung ist noch auszumachen, ob ein Kausalzusammenhang besteht oder nicht. Die Anerkennung der vorgenannten Kausalrichtungen ist hier also, in diesem Begründungszusammenhang, eine Sache der Entscheidung, die die Förderstufe bei der Verfolgung ihrer Zielsetzungen in rationaler Diskussion für sich selbst schon getroffen haben muss, d.h. es ist in dieser Untersuchung nur die Beantwortung der Frage wichtig, ob die Förderstufe ihr Ziel "Fördern statt Auslesen", bezogen auf die Anstrengungsbereitschaft, unter den von ihr vorgegebenen normativen Prämissen erreicht oder nicht.

Die isolierte Feststellung, dass die Kurszugehörigkeiten relativ stabil quasi-streaming organisiert sind, spricht in diesem

Untersuchungszusammenhang nicht gegen das Erreichen dieses Ziels. Die Fragestellung dieser Untersuchung lautet ja: Was leistet die Leistungsdifferenzierung, wie sie die Förderstufe (z.B. an dieser Schule hier) durchführt, in Bezug auf das, was die Förderstufe auch fördern will, nämlich die Anstrengungsbereitschaft? So gefragt, setzt dies die Kurszugehörigkeit als relativ invariant, gewissermassen als Kollektivvariable voraus gegenüber den individuellen Variablen Eltern-, Lehrer-Unterstützung und Anstrengungsbereitschaft. Diese Asymmetrie ist es, die die Kausalrichtung von der Kurszugehörigkeit zur Eltern-, Lehrer-Unterstützung und Anstrengungsbereitschaft rechtfertigt. Andererseits ist offensichtlich, dass die Kollektivvariable ohne den einzelnen Schüler keine Existenz hätte. Bei einem relativ stabilen quasi-streaming organisierten Kurssystem stellen die Kurszugehörigkeiten der Schüler daher, wie in der Einführung bereits ausführlich begründet, nicht nur Indikatoren der schulischen Leistungsfähigkeit eines Schülers dar, sie spiegeln zugleich auch eine durch den einzelnen Schüler nicht (oder nur wenig) veränderbare bzw. veränderte Kollektiv- oder (Kurs-)System-Variable wider. Die Frage, ob die Entstehungsprozesse für die Zusammenhänge zwischen Kurszugehörigkeit, Eltern-, Lehrer-Unterstützung und Anstrengungsbereitschaft nicht doch oder ganz allein auf der individuellen Ebene zu suchen seien, ist für unseren Test vorgegebener Zielsetzungen bzw. für unser auf Prognoseprüfung gerichtetes Untersuchungsinteresse irrelevant.

Bei der Durchführung der Untersuchung darf die Wahl der Erhebungszeitpunkte der Variablen natürlich nicht der a priori so festgelegten kausalen Ordnung widersprechen: Eltern-, Lehrer-Unterstützung und Anstrengungsbereitschaft dürfen zeitlich nicht vor, sondern müssen nach dem Stattfinden der letzten Kurszuweisung fixiert werden.

Schliesslich ist noch die Kausalrichtung zwischen der Eltern- und der Lehrer-Unterstützung a priori festzulegen, um das bei

einer hohen Korrelation unter den erklärenden Variablen drohende (Multi-)Kollinearitäts-Problem (vgl. dazu SCHNEEWEISS, 1971, S. 134 - 148) zu vermeiden. Dies erscheint schwierig angesichts der einschränkenden Bedingung, dass sowohl die Eltern- als auch die Lehrer-Unterstützung der Kurszugehörigkeit kausal nachgeordnet werden (müssen), d.h. es sieht so aus, als unterläge die Festlegung der Kausalrichtung jetzt kaum weiteren Einschränkungen mehr. Vorstellbar sind entweder nur in eine Richtung verlaufende, d.h. rekursive, oder aber in beide Richtungen verlaufende Kausalbeziehungen, Rückkopplungsbeziehungen.

Zwar betrifft der Test der Kausalhypothesen (betreffs BORCHERT & MASENDORF und Zielsetzung der Förderstufe) logisch gesehen, gleichgültig wie die Kausalrichtung zwischen der Eltern- und der Lehrer-Unterstützung auch festgelegt wird, die Kausalbeziehung zwischen ihnen nicht. Statistisch gesehen jedoch können die den Test der Kausalhypothesen direkt definierenden Modellparameter verzerrt geschätzt werden, wenn die Kausalrichtung zwischen der Eltern- und der Lehrer-Unterstützung nicht annähernd richtig spezifiziert ist. Die Konsequenz daraus ist, dass die Kausalrichtung festgelegt werden muss, und zwar so, dass sie unter Beachtung möglichst einfacher statistischer Handhabung des Gesamtmodells sowohl Plausibilitätserwägungen genügt als auch möglichst wenig Anlass zur Verzerrung der beabsichtigten Modelltests liefert.

Betrachten wir in dieser Hinsicht zunächst die Möglichkeit nur in einer Richtung verlaufender Kausalbeziehungen, so erscheint uns vornehmlich der Einfluss von der Eltern- auf die Lehrer-Unterstützung plausibel: In die Erfahrungen der Schüler mit ihren Lehrern gehen die Erfahrungen mit den Eltern mit ein. Die umgekehrte Kausalrichtung ist zwar vorstellbar, aber - in diesem Alter zumindest - vergleichsweise weniger wahrscheinlich.

Vorstellbar sind aber auch Rückkopplungsbeziehungen. Ohne eine wechselseitige Beeinflussung ganz auszuschliessen, kann man

idealisierend annehmen, dass der Rückkopplungseffekt von der Lehrer- zur Eltern-Unterstützung nach zwei Jahren recht stabil bleibender Kurszugehörigkeiten so schwach ist im Verhältnis zu der von uns bevorzugten Kausalrichtung, dass man ihn vernachlässigen kann. In einem sonst nur rekursiven Kausalmodell spielt es dann nach ANDO & FISHER (1963) für die Feststellung des Verhaltens der Variablen relativ zueinander keine wesentliche Rolle, ob man diese "schwache" Rückkopplung berücksichtigt oder nicht. Grundsätzlich sind wir natürlich der Meinung, dass Erziehung immer durch den Prozess der wechselseitigen Beeinflussung der Beteiligten bestimmt ist. Nach dem bisher Gesagten kann dies aber unbeschadet der Festlegung gelten, dass die Hauptrichtung der Beeinflussung in der von uns bevorzugten Richtung verläuft.

Unsere apriori-Festlegung der Kausalrichtung von der Eltern- zur Lehrer-Unterstützung bestimmt die Abfolge der Erhebungszeitpunkte dieser Variablen, genauer: die Darbietungsfolge der entsprechenden Messverfahren, die ihrerseits, wie noch zu zeigen sein wird, vorteilhaft dazu verwendet werden kann, systematische Messfehler aufzudecken.

Es sollen die Untersuchungshypothesen den Status von Postulaten erhalten, die eine direkte kausale Beziehung zwischen den Variablen behaupten (vgl. BLALOCK, 1969, S. 10 - 26). Um die logischen Implikationen dieser Postulate ableiten zu können, wird auch die Form der kausalen Beziehungen spezifiziert: Die im Pfaddiagramm enthaltenen Aussagen werden in mathematische Postulate übersetzt, die ein System von interdependenten Gleichungen beschreiben. Die Gleichungen werden algebraisch formuliert.

Vor der Deduktion a priori postulierter Kausalhypothesen steht der Erklärungsversuch möglicher Ereignisse. Dass wir also genaugenommen die "Entstehung" oder "Entwicklung" der Anstrengungsbereitschaft erklären, bedeutet aber nicht, dass wir das Verhalten der Variablen im Zeitablauf, das Kausalmodell also

in Form von Differentialgleichungen betrachten müssten. Denn der Aufgabenstellung der Untersuchung entsprechend sollen nicht Veränderungen oder Entwicklungen während der Förderstufenzeit beurteilt werden, sondern Entwicklungsprodukte, d.h. die aus dem Erklärungsversuch abgeleitete und zu prüfende Prognose bezieht sich auf das Ende der Förderstufenzeit, genauer: auf die Zeit nach der letzten Kurszuweisung. Für diese Zeit wird angenommen, dass die Korrelationsstruktur der Variablen annähernd ein stabiles Gleichgewicht erreicht hat, vorausgesetzt, die Bedingung eines relativ stabilen quasi-streaming organisierten Kurssystems ist erfüllt. Befindet sich die Korrelationsstruktur der Variablen in einem stabilen Gleichgewicht, dann sind im weiteren Zeitverlauf die aus der strukturellen Sicht des Zeitquerschnitts sich ergebenden Koeffizienten des simultanen Gleichungssystems identisch mit den Parametern des Prozesses (vgl. auch COLEMAN, 1968, S. 444), d.h. Querschnittsbilanzen zu späteren Zeitpunkten sollten also nicht (so bald) sehr voneinander abweichende Resultate ergeben.

2.1.6. Kontrolle des systematischen Messfehlers

Die oben angesprochene Kontrolle des systematischen Messfehlers (Formulierungsähnlichkeit der Messverfahren) ist auf analytischem Wege nur möglich, wenn in die Ueberprüfung der eigentlichen Theorie (BLALOCK, 1968 a: "general" oder "main" theory, zu deutsch auch: "Haupttheorie") zugleich die Ueberprüfung einer expliziten Messfehler-Theorie ("auxiliary" theory, zu deutsch auch: "Hilfstheorie") eingeschlossen ist. In dieser Messfehler-Theorie ist zu spezifizieren, wie unbeobachtete Variablen (auch: theoretische oder latente Variablen) und beobachtete Variablen (auch: Indikatoren) kausal zusammenhängen.

Diese Form der Fehlerkontrolle verlangt nun eine Begriffspräzisierung: Alle zu erfassenden Variablen bezeichnen wir von jetzt ab als "latente" Variable und die zu ihrer Erfassung verwendeten Verfahren als "Indikatoren". Dabei sehen wir die Indikatoren als "Wirkungen" der latenten Variablen an in dem Sinne, dass die im Indikator beobachtete Varianz von der latenten Variablen bedingt oder verursacht ist. Dementsprechend lässt sich der Betrachtung der Daten das mathematische Modell der Faktorenanalyse zugrundelegen. Bei der Faktorenanalyse von Indikatoren-Korrelationen geht man ja davon aus, dass die partielle Korrelation zwischen zwei Indikatoren dann gleich Null werden sollte, wenn eine gemeinsame Ursache, d.h. der gemeinsame Faktor oder die latente Variable konstant gehalten wird.

Die formalen Bedingungen der Faktorenanalyse stimmen mit der inhaltlichen Deutung unserer Variablen überein: Die latenten Variablen sind entweder die nicht beobachtbaren "Fähigkeiten", "Bereitschaften", also allgemeiner: die dispositionellen Eigenschaften, oder aber - wie im Falle der Eltern-/Lehrer-Unterstützung - das nicht direkt beobachtete Verhalten. Die Indikatoren sind die in der Schule und im Intelligenztest beobachteten Leistungen, die "Abbildung" des Unterstützungsverhaltens der Eltern und Lehrer im Schüler bzw. das im Fragebogen gezeigte Antwortverhalten. Nur bei Indikatoren der sozialen Herkunft (z.B. Schulbildung, Beruf, Einkommen, usw. der Eltern) jedoch könnte man sich auf den Standpunkt stellen, dass wir es nicht mit "Wirkungsindikatoren" zu tun haben, sondern umgekehrt mit Indikatoren, die die latente Variable (soziale Herkunft) "verursachen". Aber da wir aus Gründen der Anonymitätsabsicherung eine bereits vollzogene Indikatorenbündelung (zum BRSS-Indexwert, s. S. 39) zu übernehmen hatten, fällt eine Vernachlässigung dieses Standpunktes hier nicht ins Gewicht.

Im Lichte des faktorenanalytischen Modells lässt sich analytisch eine klare Unterscheidung zwischen zufälligen und systematischen Messfehlern treffen. In einer der Annahme zufälliger Messfehler entsprechenden Faktorenstruktur darf jeder Indikator nur von einem gemeinsamen Faktor bzw. latenten Variablen abhängen. Ausserdem müssen die üblichen Restriktionen der Faktorenanalyse gelten: Unabhängigkeit der latenten Variablen von den Messfehlern, Unabhängigkeit der Messfehler untereinander. Mit anderen Worten: In einem Zufallsfehler-Modell laden die einer latenten Variablen zugeordneten Indikatoren auf keinem zweiten Faktor (Modell gruppenspezifischer Faktoren). Im vorhinein wird hier also festgelegt, welche Indikatoren auf welchen Faktoren Nullladungen haben müssen. Wollen wir z.B. eine Zwei-Variablen-Beziehung: $\xi \longrightarrow \eta$ überprüfen, mit drei verschiedenen Indikatoren für jede Variable, dann sieht die Faktorenstruktur, die bei Fehlen systematischer Messfehler gelten bzw. auf die Daten passen müsste, im Pfaddiagramm so aus:

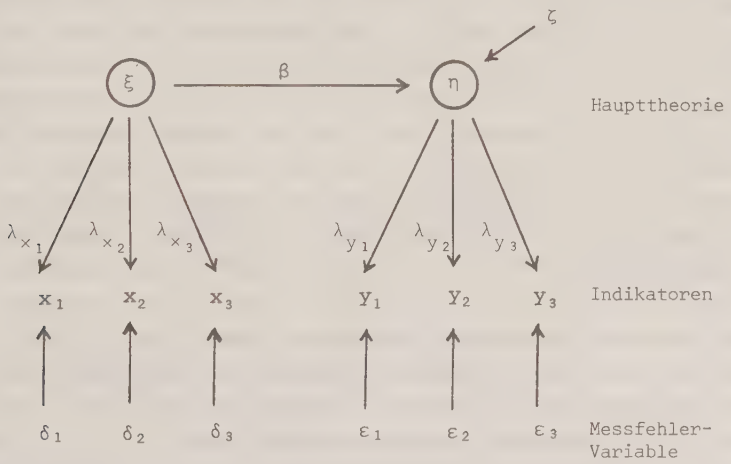

Abb. 7: Beispiel eines Zufallsfehler-Modells: zwei latente Variable (ξ und η) mit je drei Indikatoren. Wobei: ζ = latente Residualvariable; β = Koeffizient der Regression von η auf ξ; λ = Faktorenladungen

Die im Zufallsfehler-Modell postulierte Kausalstruktur fällt zwar unter die allgemeine Definition des faktorenanalytischen Modells, doch ist es nicht möglich, die hier aufgeworfene Fragestellung mit Hilfe der klassischen explorativen Faktorenanalyse (vgl. HARMAN, 1967) zu behandeln. Es kommt hier ja auch nicht darauf an, irgendwelche die Daten gut reproduzierende Faktorenstrukturen zu suchen oder zu finden (explorativ Vorgehen), sondern darauf, kausale Hypothesen zu testen, also zu sehen, ob mit bestimmten Restriktionen (z.B. Gleichheits- oder Null-Restriktionen von Faktoren-Ladungen oder -Korrelationen) vorgegebene Faktorenstruktur sich an empirischem Datenmaterial "bewähren" (konfirmatives Vorgehen).

Ein Schätzverfahren, das die in diesem Zusammenhang vorgegebenen Restriktionen angemessen berücksichtigt, ist die konfirmative Faktorenanalyse von JOERESKOG (1969). Die bei der explorativen Faktorenanalyse allein aus mathematischen Gründen der eindeutigen Lösbarkeit des Gleichungssystems verwendete Diagonalisierungsrestriktion, eine im Lichte der Daten oder psychologischen Theorie möglicherweise oft nicht zu rechtfertigende Modellannahme, wird bei der konfirmativen Faktorenanalyse nicht benötigt. Das Unbestimmtheitsproblem kann hier in begründeten Fällen auf einfachere, weniger annahmenbelastete Weise dadurch gelöst werden, dass bei der Schätzung der Parameter (mit dem Computerprogramm COFAMM = Confirmatory Factor Analysis with Model Modification, von SOERBOM & JOERESKOG, 1976) die durch die zu testende Hypothese mitgegebenen apriori-Kenntnisse über das Kausalmodell mitbeachtet werden. Durch theoriegeleitete apriori-Restriktionen einiger Parameter (z.B. Null-Ladungen einiger Indikatoren) kann so genügend Information zur Identifikation des Kausalmodells bereitgestellt werden.

Für das in Abb. 7 betrachtete Modell allein reicht die Analyse mit COFAMM aus. Denn bei Zwei-Variablen-Modellen macht

es rein rechnerisch keinen Unterschied, ob wir die Analyse vom Standpunkt der obliquen konfirmativen Faktorenanalyse oder vom Standpunkt des Einzelgleichungsmodells der einfachen Regression her betrachten: Die von COFAMM errechnete Faktorenkorrelation ist hier mit dem Koeffizienten der standardisierten Faktorenregression identisch.

In unserer Untersuchung jedoch gehen wir von mehreren (und nicht nur von einer einzigen) erklärenden latenten Variablen aus, welche untereinander kausal zusammenhängen. Wenn wir deshalb die postulierte Modellstruktur durch so viele Gleichungen beschreiben müssen wie wir endogene (d.h. innerhalb des Modells zu erklärende) latente Variablen haben, dann benötigen wir ein Analyseverfahren, das bei der Schätzung der Strukturparameter auch noch die Tatsache berücksichtigt, dass jede Gleichung in ein Gleichungssystem gehört.

Ein zur Analyse solcher Strukturgleichungs-Modelle geeignetes Verfahren wurde ebenfalls von JOERESKOG (1970) vorgeschlagen. Zur Anwendung dieses Verfahrens wurde von JOERESKOG & van THILLO (1972) das Computerprogramm LISREL zur Verfügung gestellt (LISREL = Analysis of Linear Structural Relationships; wobei der Name "LISREL" im folgenden sowohl für eine bestimmte Klasse von Kausalmodellen bzw. Schätzverfahren steht als auch für das zur Schätzung der Modellparameter verwendete Computerprogramm). Hier werden, wie auch bei COFAMM, die Modellparameter nach dem Maximum-Likelihood-Prinzip geschätzt. Die Güte der Anpassung von Modell und Daten wird statistisch über einen χ^2-Test beurteilt.

Nicht vorgesehene systematische Messfehler verschlechtern nun die Anpassungsgüte des Modells der zufälligen Messfehler. Da LISREL alle Strukturgleichungen des Modells gleichzeitig bestimmt und das System von Haupt- und Messfehler-Theorie zusammen als Ganzes schätzt ("Schätzverfahren bei voller Information"), sind nicht-experimentelle apriori-Kontrollen von sys-

tematischen Messfehlern nur einzurichten, wenn wir explizit Hypothesen postulieren, in welcher Richtung diese Messfehler wirken.

In diesem Zusammenhang lautet die technische Definition des systematischen Messfehlers so: Ein systematischer Messfehler liegt vor, wenn die zusätzliche Nullrestriktion, die das Modell mit dem systematischen Messfehler zum Modell der zufälligen Messfehler verschärft (also mindestens einen freien Parameter weniger hat), eine gravierende Verschlechterung in der Modellanpassung nach sich zieht.

Wie soll nun die Wirkungsrichtung einer Antwortverzerrung durch Formulierungsähnlichkeit festgelegt werden? Lässt man die Items formulierungsähnlicher Messverfahren ungemischt, dann hängt definitionsgemäss die Wirkung der Antwortverzerrung von der Darbietungsfolge der Verfahren ab. Die Wirkungsrichtung sollten wir so zu kanalisieren versuchen, dass die sich ergebende Kombination von Messfehler- und Haupttheorie für LISREL möglichst einfach handhabbar bleibt. Dies ist der Fall, wenn wir die zu erwartende Wirkung des systematischen Messfehlers mit der zwischen den latenten Variablen a priori festgelegten Wirkungsrichtung gleichgerichtet halten. Dementsprechend (Elternunterstützung "kausal vor" Lehrerunterstützung) entscheiden wir uns für die Darbietungsfolge: zuerst die Eltern-Unterstützungs-Skalen, dann die Lehrer-Unterstützungs-Skala. Liegt bei dieser Darbietungsfolge eine Beeinflussung durch Formulierungsähnlichkeit vor, dann bedeutet das, dass Schüler mit einer hohen (bzw. niedrigen) Eltern-Unterstützung auch hohe (bzw. niedrige) Werte auf der Lehrer-Unterstützungs-Skala haben, unabhängig davon, welche Lehrer-Unterstützung sie tatsächlich erleben.

Um die Erfassung der Lehrer-Unterstützung nicht völlig mit dem vermuteten Messfehler zu konfundieren, wurde vorsichtshalber ein weiterer Indikator hereingenommen: die Lehrer-Sympathie-Skala von BRUSTEN & HURRELMANN (s. oben). Der im Ansatz von

LISREL überprüfte systematische Messfehler liegt nun vor, wenn der formulierungsähnliche Indikator der Lehrer-Unterstützung nicht spezifisch misst, sondern auch die ihm theoretisch nicht zugeordnete Variable "Eltern-Unterstützung" erfasst, also im Unterschied zum weniger formulierungsähnlichen Indikator auf zwei latenten Variablen bzw. Faktoren Ladungen ungleich Null hat. Im Pfaddiagramm sieht dieser Sachverhalt (Modellausschnitt) so aus:

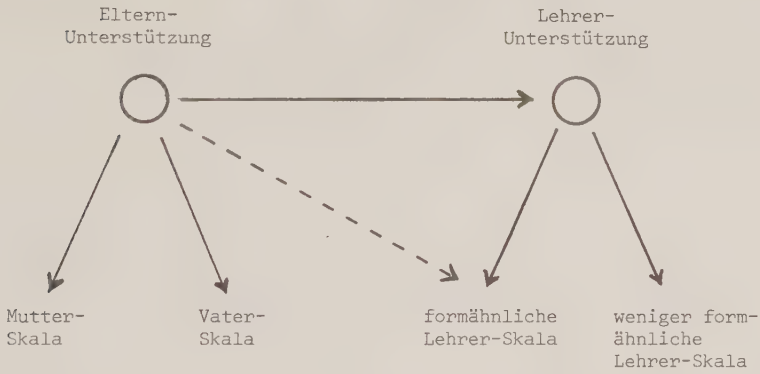

Abb. 8: Modellausschnitt zur Veranschaulichung des postulierten systematischen Messfehlers (gestrichelt): Antwortverzerrung durch Form(ulierungs)ähnlichkeit

Die Falsifikation des Zufallsfehler-Modells führt zur Ablehnung bestimmter Indikatoren bzw. Indikatorenzusammenstellungen zur Prüfung der eigentlichen Theorie. Muss man in unserem Zusammenhang realistischerweise den systematischen Messfehler unterstellen, so wird zur Erfassung der Lehrer-Unterstützung auf das formulierungsähnliche Messverfahren verzichtet und stattdessen der Zusatzindikator verwendet.

2.1.7. Kontrolle der Repräsentativität von Messungen

Bisher war der primäre Grund für das Arbeiten mit multiplen Indikatoren die Kontrolle von Fehlern, die durch forschungstechnische Eigentümlichkeiten verursacht werden. Wenn keine systematischen Messfehler vorliegen, lassen sich Indikatoren aber auch zur Kontrolle der Generalisierbarkeit eines Ergebnisses auf andere Messinstrumente verwenden: Mit unabhängigen Mehrfachmessungen von Variablen kann die Gefahr mangelnder Korrespondenz zwischen zu prüfender Theorie und Messoperationen reduziert werden. Unter diesem Gesichtspunkt kann man sich fragen, ob der zunächst zur Messfehlerkontrolle zugezogene Indikator der Lehrer-Unterstützung auch bei der Beurteilung der eigentlichen Theorie hilfreich sein kann. Wir müssen dann überlegen, unter welchen Bedingungen dies geschehen kann.

Der Einsatz des Zusatzindikators soll bei der Ueberprüfung der eigentlichen Theorie nur unter dem Nachweis zu rechtfertigen sein, dass dieser dem zuerst ausgewählten Indikator in etwa gleichwertig ist. Diesen Gleichwertigkeitsnachweis wollen wir auch von den Indikatoren der Eltern-Unterstützung und der schulischen Leistungsfähigkeit verlangen. Im analytischen Rahmen von LISREL wollen wir die Indikatoren einer latenten Variablen dann als "gleichwertig", "gleich gut", "auswechselbar" oder "parallel" bezeichnen, wenn sie - ihre Standardisierung vorausgesetzt - gleiche interne Reliabilitäten bzw. gleiche Messfehler-Varianzen und gleiche externe Validitäten (in Bezug auf die Variablen der Theorie, in der sie eingebettet sind) haben. Diese Festlegung besagt,

1. dass die Interkorrelationen jeweils paralleler Indikatoren gleich sein müssen, und
2. dass jeweils parallele Indikatoren mit jedem beliebigen Indikator des Modells gleich hoch korrelieren müssen.

Da wir die latenten Variablen des Modells höchstens zweifach messen, ist die erste Bedingung automatisch erfüllt, d.h. bei nur zwei Indikatoren je latenter Variable sind die restriktiven Bedingungen für die Parallelität leichter zu erfüllen als bei drei oder mehr Indikatoren je latenter Variable.

Nur die Gleichheit der in der zweiten Bedingung genannten Korrelationen ist hier von Belang. In einem Strukturgleichungs-Modell, dessen Parameter unter Verwertung aller Informationen der Kausalstruktur geschätzt werden (durch LISREL), hat eine Ungleichheit dieser Korrelationen oder externen Validitäten "Rückwirkungen" auf die Höhe der Faktorenladungen sonst äquivalenter Indikatoren, d.h. es ergeben sich uneinheitlichere, von der Gleichheitsrestriktion stärker abweichende Ladungswerte als wenn man diese Indikatoren abgeblockt von den Indikatoren anderer latenter Variablen des Modells faktorenanalysiert hätte.

Aus der strukturellen Mitbedingtheit der Höhe der Faktorenladungen folgt, dass eine eindeutige Interpretation einer latenten Variablen nicht möglich ist, wenn die ihr theoretisch zugeordneten Indikatoren unterschiedlich enge Beziehungen zu anderen Indikatoren des Modells haben. Dieses Ergebnis deckt sich mit der Vorstellung, dass eine latente Variable nicht allein durch ihre Indikatoren, sondern partiell auch durch ihre Verknüpfung mit anderen latenten Variablen definiert wird.

2.1.8. Hypothesen dieser Untersuchung, prüfbar und vollständig: LISREL-Modellspezifikation

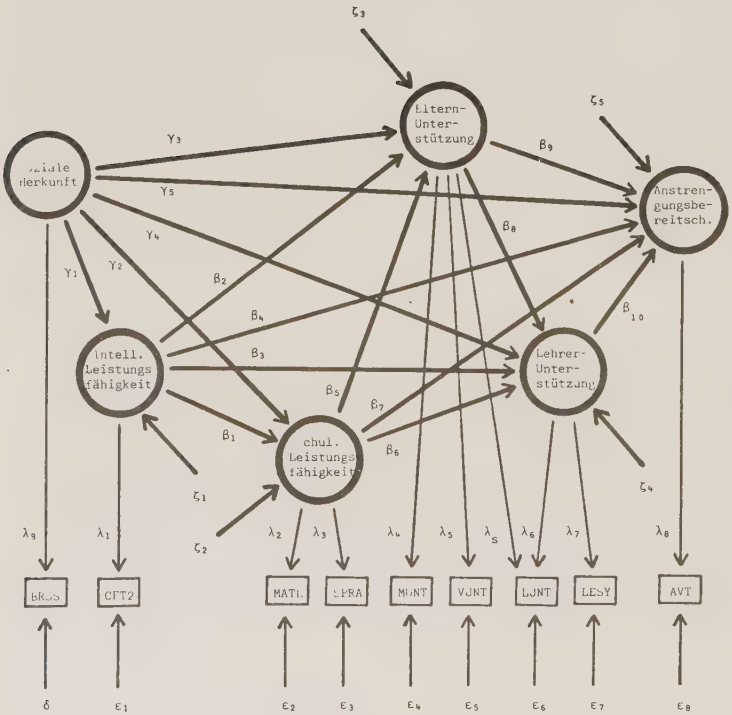

Abb. 9: Vollständig rekursives Strukturgleichungsmodell zur Erklärung der Anstrengungsbereitschaft. Die Richtungen der Kausalbeziehungen sind positiv.
Kreise: latente Variable
Rechtecke: Indikatoren, wobei bedeuten: BRSS = Verfahren zur Messung des sozialen Status von BAUER, CFT2 = Grundintelligenztest-Skala 2 von WEISS, MATH und SPRA = Kurszugehörigkeiten in Mathematik und Sprachen am Ende der Förderstufe, MUNT und VUNT = Marburger Unterstützungsskalen: Mutter- und Vaterversion, LUNT = Lehrer-Unterstützungsskala von HEINRICH, LESY = Skala zur Erfassung der Einschätzung der Lehrersympathie von BRUSTEN und HURRELMANN, AVT = Anstrengungsvermeidungstest von ROLLETT und BARTRAM

Es sei:

Indikatoren

BRSS = x
CFT2 = y_1
MATH = y_2
SPRA = y_3
MUNT = y_4
VUNT = y_5
LUNT = y_6
LESY = y_7
AVT = y_8

Latente Variablen

Soziale Herkunft = ξ
Intellektuelle Leistungsfähigkeit = η_1
Schulische Leistungsfähigkeit = η_2
Eltern-Unterstützung = η_3
Lehrer-Unterstützung = η_4
Anstrengungsbereitschaft = η_5

LISREL-Spezifikation des Messmodells von Abb. 9, in algebraischer Form:

$$x = \lambda_9 \xi + \delta$$
$$y_1 = \lambda_1 \eta_1 + \varepsilon_1$$
$$y_2 = \lambda_2 \eta_2 + \varepsilon_2$$
$$y_3 = \lambda_3 \eta_2 + \varepsilon_3$$
$$y_4 = \lambda_4 \eta_3 + \varepsilon_4$$
$$y_5 = \lambda_5 \eta_3 + \varepsilon_5$$
$$y_6 = \lambda_s \eta_3 + \lambda_6 \eta_4 + \varepsilon_6$$
$$y_7 = \lambda_7 \eta_4 + \varepsilon_7$$
$$y_8 = \lambda_8 \eta_5 + \varepsilon_8$$

Der Einfachheit halber sind die Indikatoren vom jeweiligen Mittelwert aus gemessen. Dadurch entfallen die Lokalisationsparameter μ und ν in den Gleichungen.

In Matrix-Darstellung:

$$\mathbf{x} = \lambda_9 \xi + \delta$$

und

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & \lambda_3 & 0 & 0 & 0 \\ 0 & 0 & \lambda_4 & 0 & 0 \\ 0 & 0 & \lambda_5 & 0 & 0 \\ 0 & 0 & \lambda_s & \lambda_6 & 0 \\ 0 & 0 & 0 & \lambda_7 & 0 \\ 0 & 0 & 0 & 0 & \lambda_8 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \end{bmatrix}$$

LISREL-Spezifikation des Strukturmodells von Abb. 9:
Strukturgleichungen, in algebraischer Form:

$$\eta_1 = \gamma_1 \xi + \zeta_1$$
$$\eta_2 = \gamma_2 \xi - \beta_1 \eta_1 + \zeta_2$$
$$\eta_3 = \gamma_3 \xi - \beta_2 \eta_1 - \beta_5 \eta_2 + \zeta_3$$
$$\eta_4 = \gamma_4 \xi - \beta_3 \eta_1 - \beta_6 \eta_2 - \beta_8 \eta_3 + \zeta_4$$
$$\eta_5 = \gamma_5 \xi - \beta_4 \eta_1 - \beta_7 \eta_2 - \beta_9 \eta_3 - \beta_{10} \eta_4 + \zeta_5$$

oder in Matrix-Darstellung:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \beta_1 & 1 & 0 & 0 & 0 \\ \beta_2 & \beta_5 & 1 & 0 & 0 \\ \beta_3 & \beta_6 & \beta_8 & 1 & 0 \\ \beta_4 & \beta_7 & \beta_9 & \beta_{10} & 1 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \\ \eta_5 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \end{bmatrix} \xi + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \\ \zeta_5 \end{bmatrix}$$

Der Einfachheit halber sind die latenten Variablen hier standardisierte Variablen.

Hierbei sind:

ξ : exogene latente Variable
η_i : endogene latente Variable
ζ_i : Residuen der latenten Variablen ("Fehler in den Strukturgleichungen")
x : Indikator der exogenen latenten Variable
y_i : Indikatoren der endogenen latenten Variablen
δ : Messfehler-Variable des x-Indikators
ε_i : Messfehler-Variable der y-Indikatoren
γ_i : Strukturparameter: Koeffizienten der Regression der endogenen latenten Variablen auf die exogene latente Variable
β_i : Strukturparameter: Koeffizienten der Regression der endogenen latenten Variablen auf andere endogene latente Variable
λ_i : Messmodellparameter: Koeffizienten der Regression der Indikatoren auf die latenten Variablen (in der Terminologie der common-factor-Theorie die "Faktorenladungen")

Die oben aufgeführten Gleichungssysteme spiegeln teilweise die Annahmen von LISREL wider, ohne die die Modellparameter nicht geschätzt werden können. Es ist sinnvoll, die wichtigsten Modellannahmen bzw. Restriktionen (auch: "vereinfachende" Bedingungen, die für die "Schliessung des Systems" sorgen, s. S. 29) explizit zu machen:

1. $E(\eta) = E(\xi) = E(\zeta) = E(\varepsilon) = E(\delta) = 0$

   d.h. die Erwartungswerte der latenten Variablen und der Messfehler-Variablen sind Null.

2. $E(\xi\zeta') = 0$
   d.h. die Residuen der (endogenen) latenten Variablen korrelieren nicht mit der exogenen latenten Variablen.

3. Die Matrix der β-Werte, B, ist nicht-singulär,

   d.h. keine der endogenen latenten Variablen ist von einer anderen endogenen latenten Variablen linear abhängig.

4. $E(\varepsilon\delta') = 0$; $E(\varepsilon\varepsilon') = \Theta_\varepsilon^2$; $E(\delta\delta') = \Theta_\delta^2$, wobei:

   Θ_ε^2 und Θ_δ^2 : Diagonalmatrizen

   d.h. die Messfehler-Variablen korrelieren nicht untereinander.

5. $E(\eta\varepsilon') = E(\xi\delta') = E(\eta\delta') = E(\xi\varepsilon') = E(\zeta\varepsilon') = E(\zeta\delta') = 0$

   d.h. die Messfehler-Variablen korrelieren nicht mit den latenten Variablen.

Aus den genannten Annahmen lässt sich nach JOERESKOG (1970) bzw. JOERESKOG & van THILLO (1972, S. 3) folgende Varianz-Kovarianz-Matrix Σ $[(8 + 1) \times (8 + 1)]$ der Indikatoren $z = (y', x')'$ herleiten:

$$\Sigma = \left[\begin{array}{c|c} \Lambda_y(B^{-1}\Gamma\Phi\Gamma'B'^{-1} + B^{-1}\Psi B'^{-1})\Lambda_y' + \Theta_\varepsilon^2 & \Lambda_y B^{-1}\Gamma\Phi\Lambda_x' \\ \hline \Lambda_x\Phi\Gamma'B'^{-1}\Lambda_y' & \Lambda_x\Phi\Lambda_x' + \Theta_\delta^2 \end{array}\right]$$

wobei:

Λ_x bzw. Λ_y : Matrizen der Regressionen (λ) der x-Indikatoren auf ξ bzw. der y-Indikatoren auf η. Hier: $\Lambda_x(1x1)$ und $\Lambda_y(8x5)$

B : Matrix der Regressionen (β) der endogenen latenten Variablen (η) auf andere endogene latente Variable. Hier: B(5x5)

Γ : Matrix der Regressionen (γ) der endogenen latenten Variablen (η) auf die exogenen latenten Variablen (ξ). Hier: Γ(5x1)

Φ : Varianz-Kovarianz-Matrix der exogenen latenten Variablen (ξ). Hier: $\Phi(1x1) = Var(\xi)$

Ψ : Varianz-Kovarianz-Matrix der Residuen (ζ) der latenten Variablen. Hier: $\Psi(5x5) = diag(\psi_{ii})$, $\psi_{ii} = Var(\zeta_i)$, i = 1,2,3,4,5 - d.h. eine Diagonalmatrix mit den in η_i durch das Modell nicht-erklärten Varianzen.

Θ^2_δ bzw. Θ^2_ε : Diagonalmatrizen der Messfehler-Variablen, d.h. nur die Messfehler-Varianzen der x- bzw. y-Indikatoren. Hier: $\Theta^2_\delta(1x1) = diag(\theta^2_\delta)$, $\theta^2_\delta = Var(\delta)$ bzw. $\Theta^2_\varepsilon(8x8) = diag(\theta^2_{\varepsilon i})$, $\theta^2_{\varepsilon i} = Var(\varepsilon_i)$, i = 1, 2,3,4,5,6,7,8.

Die Matrix Σ besteht aus folgenden vier Teilmatrizen:

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{yy}(8x8) & \Sigma_{yx}(8x1) \\ \hline \Sigma_{xy}(1x8) & \Sigma_{xx}(1x1) \end{array}\right]$$

wobei:

Σ : Varianz-Kovarianz-Matrix sämtlicher Indikatoren
Σ_{yy} : Varianz-Kovarianz-Matrix aller y-Indikatoren
Σ_{xx} : Varianz-Kovarianz-Matrix aller x-Indikatoren
Σ_{yx} : Matrix der Kovarianzen der x- mit den y-Indikatoren
Σ_{xy} : zu Σ_{yx} symmetrisch

Die Elemente von Σ sind Funktionen der Elemente von Λ_y, Λ_x, B, Γ, Φ, Ψ, Θ^2_δ und Θ^2_ε. Im Anwendungsfall sind einige dieser Elemente a priori fixiert. Im LISREL-Verfahren werden unter der Voraussetzung, dass die Indikatoren multinormal verteilt sind, die restlichen freien Parameter bei Berücksichtigung der apriori-Restriktionen so geschätzt, dass die Likelihood der beobachteten Varianz-Kovarianz-Matrix S maximiert wird. Die geschätzten Parameter sind dann Maximum-Likelihood-Koeffizienten, welche effizient in grossen Stichproben sind.

Entsprechend unseren theoretischen Ausführungen sollen die Daten zunächst unter den folgenden Hypothesen H_1 - H_5 nacheinander analysiert werden:

H_1 : entspricht dem oben bereits spezifizierten Struktur- und Messmodell von Abb. 9

H_2 : $\lambda_s = 0$

H_3 : $\lambda_s = 0$, $\lambda_2 = \lambda_3$, $\lambda_4 = \lambda_5$, $\lambda_6 = \lambda_7$,
$Var(\varepsilon_2) = Var(\varepsilon_3)$, $Var(\varepsilon_4) = Var(\varepsilon_5)$, $Var(\varepsilon_6) = Var(\varepsilon_7)$.

H_4 : $\lambda_s = 0$, $\lambda_2 = \lambda_3$, $\lambda_4 = \lambda_5$, $\lambda_6 = \lambda_7$,
$Var(\varepsilon_2) = Var(\varepsilon_3)$, $Var(\varepsilon_4) = Var(\varepsilon_5)$, $Var(\varepsilon_6) = Var(\varepsilon_7)$,
$\gamma_3 = 0$, $\gamma_4 = 0$, $\gamma_5 = 0$, $\beta_9 = 0$.

H_5 : $\lambda_s = 0$, $\lambda_2 = \lambda_3$, $\lambda_4 = \lambda_5$, $\lambda_6 = \lambda_7$,
$Var(\varepsilon_2) = Var(\varepsilon_3)$, $Var(\varepsilon_4) = Var(\varepsilon_5)$, $Var(\varepsilon_6) = Var(\varepsilon_7)$,
$\gamma_3 = 0$, $\gamma_4 = 0$, $\gamma_5 = 0$, $\beta_9 = 0$,
$\beta_2 = 0$, $\beta_3 = 0$, $\beta_4 = 0$.

Dabei bedeuten:

H_1 : Hypothese des systematischen Messfehlers

H_2 : Hypothese zufälliger Messfehler

H_3 : Hypothese paralleler Indikatoren

H_4 : Erste Alternative zur Vermittlungshypothese: keine direkte Vermittlung durch Eltern- und Lehrer-Unterstützung

H_5 : Zweite Alternative zur Vermittlungshypothese, als Spezialfall der ersten Alternative: entscheidende Vermittlung durch die Kurszugehörigkeiten

Ergänzend ist zu bemerken: Auf das Strukturmodell bezogen repräsentieren H_1, H_2 und H_3 auch die Hypothese der Vermittlung durch Eltern- und Lehrer-Unterstützung.

Dann ist beabsichtigt, fünf Hypothesen gegen zunehmend weniger allgemeine Alternativhypothesen zu testen, und zwar:

1. H_2 gegen H_1 : Test der Hypothese zufälliger Messfehler
2. H_3 gegen H_2 : Test der Hypothese paralleler Indikatoren
3. H_4 gegen H_3 : Test der ersten Alternative zur Vermittlungshypothese
4. H_5 gegen H_3 : Test der zweiten Alternative zur Vermittlungshypothese
5. Zusatzrestriktion $\beta_7 = 0$ gegen H_5: Test einer Zielsetzung der Förderstufe

Zunächst einige mehr technische Bemerkungen dazu:

Vor diesen Tests wird in jeder Geschlechtergruppe jede Hypothese separat gegen die allgemeinste Alternativhypothese getestet, dass Σ nicht restringiert ist. Bei der Gegenüberstellung der Hypothesen ist zu beachten, dass die Nullhypothese stets einen geringeren Allgemeinheitsgrad hat als die jeweilige Alternativhypothese, d.h. die Restriktionen in den fünf zu testenden Nullhypothesen sind hierarchisch angeordnet: Zuerst gilt $\lambda_s = 0$, sodann $\lambda_s = 0$ zusätzlich mit $\lambda_2 = \lambda_3$, $\lambda_4 = \lambda_5$, $\lambda_6 = \lambda_7$, usw. Die Durchführung aller hier vorgeschlagenen Tests setzt daher die Nicht-Verwerfung der zugehörigen Alternativhypothese voraus. Im Falle der Verwerfung z.B. von H_2 muss das Messmodell neu spezifiziert werden und die Testabfolge ändert sich dementsprechend.

Die genannte Testabfolge bedarf weiterer, mehr inhaltlicher Erläuterungen:

1. Zu den ersten vier Tests: Wie bereits aus der Einleitung der Arbeit hervorgeht, wird die Modellbildung dieser Untersuchung zugespitzt auf die Alternative: "keine entscheidende Vermittlung durch die Kurszugehörigkeit" (H_1 oder H_2 oder H_3) versus "entscheidende Vermittlung durch die Kurs-

zugehörigkeit" (H_5). Die "zwischen" der Alternative H_1 oder H_5 liegenden Kausalhypothesen H_2, H_3 und H_4 sind sukzessive apriori-Verschärfungen von H_1 (d.h. immer mehr Nullrestriktionen) und sollen ein sequentielles Testen ermöglichen. Prinzipiell kann natürlich auch direkt getestet werden, nämlich gleich H_5 gegen H_1. Sequentiell soll hier aber von vornherein getestet werden wegen der Messfehler-Hypothesen und weil bereits H_4 eine zur Vermittlungshypothese BORCHERT & MASENDORFs konkurrierende Hypothese ist und nicht erst H_5.

In den Tests "H_4 gegen H_3" und "H_5 gegen H_3" repräsentiert H_3 die Hypothese einer Vermittlung (durch Eltern- und Lehrer-Unterstützung) im weiteren wie engeren Sinne zugleich (siehe dazu S. 7 in dieser Arbeit). Da zweifelhaft ist, ob BORCHERT & MASENDORF mit "kausaler Vermittlung" die im üblich engeren Sinne einfache kausale Kette (d.h. hier: $\gamma_5 = 0$) meinen, muss eine durchaus denkmögliche Begriffserweiterung von "Vermittlung" (d.h. hier: γ_5 als freier Parameter) mit in Betracht gezogen werden, zumal Vagheit in der Begriffsbestimmung ja noch nicht Willkür in der Entscheidung über Falsifikationsmöglichkeiten impliziert: Um uns nicht von vornherein dem Risiko des Vorwurfs einer falschen Auslegung einer unscharf formulierten Hypothese auszusetzen, testen wir H_4 und H_5 nicht gegen $\gamma_5 = 0$, sondern gegen eine Hypothese (H_3), in der alle Strukturparameter freie Parameter sind. Im Falle der Verwerfung von H_3 hätte dieser Test dann einen "doppelgleisigen" Effekt: Er lässt dieselbe klare Entscheidung gegen jede der in der Untersuchung BORCHERT/MASENDORFs implizierten Möglichkeiten von Vermittlung zu. Wegen der Hierarchie der Restriktionen absorbiert die Falsifikation von H_3 die engere Auffassung von Vermittlung: Die Falsifikation der weiteren Auffassung von Vermittlung falsifiziert hier die engere Auffassung von Vermittlung mit.

2. Zum letzten, dem fünften Test: Wie aus der Einleitung der Arbeit weiter hervorgeht, soll hier neben der Vermittlungshypothese BORCHERT & MASENDORFs auch eine Zielsetzung der Förderstufe getestet werden. Mit der Zuspitzung der Modellbildung auf die Alternative "keine entscheidende Vermittlung durch die Kurszugehörigkeit" versus "entscheidende Vermittlung durch die Kurszugehörigkeit" wird zugleich auch die Voraussetzung für einen kausalanalytisch präzisierten Test der Zielsetzung "Fördern statt Auslesen", bezogen auf die Anstrengungsbereitschaft, geschaffen.

   Im Rahmen des vorgegebenen Kausalmodells entspricht der Test der Zusatzrestriktion $\beta_7 = 0$ gegen H_5 einem Entscheidungstest zwischen den Möglichkeiten (vereinfacht formuliert) "Fördern statt Auslesen" versus "entscheidende Vermittlung durch die Kurszugehörigkeit". Aufgrund der kausalstrukturellen Sichtweise ist dabei zu beachten, dass die im Ausdruck "Fördern statt Auslesen" enthaltenen zwei Begriffe "Förderung" und "Auslese" hier nicht separat betrachtet werden (können). Definiert wird der gesamte Ausdruck "Fördern statt Auslesen" dadurch, dass zusätzlich zu den Nullrestriktionen von H_5 die Nullrestriktion $\beta_7 = 0$ (β_7 betrifft die Kausalbeziehung zwischen Kurszugehörigkeit und Anstrengungsbereitschaft) a priori postuliert wird. Mit dem isolierten Nullsetzen z.B. von β_7 und dem anschliessenden Testen nur dieser Restriktion z.B. gegen das vollständig rekursive Strukturmodell (H_1 oder H_2 oder H_3) ist es daher nicht getan. Die einzelne kausale Beziehung, werde sie nun als vorhanden oder nicht vorhanden postuliert, wird mitdefiniert durch den kausalen Gesamtzusammenhang, in den sie eingebettet ist.

Zusammenfassend: Die Kombination der drei ersten in der Einleitung angeführten Teilmodelle (Abb. 1, 2 und 3; Abb. 2 inklusive der erweiterten Definition von Vermittlung) zu einem komplexen Kausalmodell auf der einen Seite und auf der anderen Seite:

die aufgrund alternativer sozialpsychologischer Theorien (z.B. Kausalattributions-Theorie, Theorien des sozialen Vergleichs, der sozialen Erwartung) nahegelegte Modifikation dieses komplexen Kausalmodells führen zu der Alternative: "keine entscheidende Vermittlung durch die Kurszugehörigkeit" versus "entscheidende Vermittlung durch die Kurszugehörigkeit". Die prüfbar formulierten Kausalhypothesen H_1 (oder H_2 oder H_3) und H_5 sollen den Entscheidungstest zwischen den genannten zwei Möglichkeiten erlauben, d.h. einen fairen Test von BORCHERT & MASENDORFs Kausalhypothese ermöglichen und zugleich die Grundlage für einen kausalanalytisch präzisierten Test einer Förderstufen-Zielsetzung liefern.

Diese Aufgabenstellungen implizieren nicht nur eine bestimmte Modellhierarchie, sie legen weitgehend auch die Kausalrichtung a priori, als in derselben Untersuchung zunächst nicht hinterfragbar fest. Verantwortungsbewusstes Arbeiten mit LISREL setzt darüber hinaus voraus, dass man explizit zu prüfende Kausalhypothesen in bestimmte Strukturgleichungen überträgt, also formalisiert. Das heisst mit anderen Worten: Die Herleitung der Fragestellung der Untersuchung ist vor der LISREL-Modellspezifikation noch nicht abgeschlossen. Dementsprechend wurden in der Herleitung der Fragestellung keine am selben Datenmaterial explizit zu prüfenden Alternativhypothesen genannt.

Bei komplexen Kausalmodellen versteht es sich von selbst, dass darin enthaltene einzelne Kausalhypothesen (z.B. "die Kurszugehörigkeit wirkt auf die Anstrengungsbereitschaft") nicht isoliert vom kausalen Gesamtzusammenhang betrachtet werden dürfen. Wenn in der Arbeit im Entstehungs- und Begründungs-Zusammenhang des Kausalmodells, also im Stadium der Modellbildung, auf einzelne Kausalhypothesen näher eingegangen wird, so ist das selbstverständlich immer im Kontext des bereits fertig (vor-)gedachten gesamten Modells gesehen. Die in der Modellbildung zunächst einzeln behandelte Kausalhypothese gilt immer nur im mitgegebenen Variablen-Kontext. Analog ist eine empi-

risch (vorläufig) gestützte direkte kausale Beziehung zwischen zwei Variablen nur "direkt" im Hinblick auf die im betrachteten Modell explizierten Variablen.

2.2. Analyseverfahren

2.2.1. Identifizierbarkeit, Schätzen und Testen

Die Modellparameter werden mit dem Computerprogramm LISREL von JOERESKOG & van THILLO (1972) geschätzt. Vor der Schätzung der Parameter ist zu prüfen, ob diese auch identifizierbar sind.

Zur Identifizierbarkeit: Anhand von GOLDBERGERs Version des Abzählkriteriums (1964, S. 316) kann man sich davon überzeugen, dass jedes der hier zu analysierenden Modelle als ganzes identifiziert ist (den Begriff "Modell" kann man als dem Begriff "formalisierte Hypothese" gleichbedeutend ansehen). Allerdings garantiert dieses Kriterium dafür nicht hinreichend. Im übrigen fehlt auch ein allgemeines Verfahren, das es gestattete, jedes Modell multipler Indikatoren eindeutig auf Identifizierbarkeit zu prüfen (vgl. WILEY, 1973), wie das FISHER (1966) mit seinen Theoremen für Modelle mit direkt gemessenen Variablen erreicht. Ein zumindest praktisch ausreichendes Kriterium für den Identifikationsstatus eines Modells multipler Indikatoren sehen JOERESKOG & SOERBOM (1976, S. 8-9) darin, ob die zur Berechnung der Standardschätzfehler verwendete Informationsmatrix positiv definit ist oder nicht: Meldet das LISREL-Programm keine Singularität der Informationsmatrix, dann kann man, von extremen Sonderfällen abgesehen, davon ausgehen, dass das Modell identifiziert ist.

Zum Schätzen und Testen: Von den im LISREL-Modell getroffenen Annahmen ausgehend besteht das Schätzproblem im wesentlichen darin, die durch das Modell implizierte Matrix Σ möglichst gut an die beobachtete Matrix S anzupassen. Im LISREL-Verfahren vergleicht die Likelihood-Ratio-Teststatistik die Nullhypothese mit der allgemeinsten Alternativhypothese, dass die positiv definite Matrix Σ ohne Restriktionen ist. In grossen Stichproben hat diese Statistik eine χ^2-Verteilung mit Freiheitsgraden gleich der Differenz zwischen der Anzahl der Varianzen und Kovarianzen und der Anzahl der unter dem spezifizierten Modell zu schätzenden unabhängigen Parametern. Entsprechend ist auch der zum χ^2-Wert gehörende P-Wert zu interpretieren: P ist definiert als die Wahrscheinlichkeit, einen grösseren χ^2-Wert als den tatsächlich erhaltenen zu bekommen unter der Voraussetzung, dass das postulierte Modell (die Nullhypothese) wahr ist. Daher entsprechen kleine Werte von P einer (statistisch gesehen) schlechten Modellanpassung und grosse Werte von P einer guten Anpassung.

Von der Güte der Modellanpassung muss man sich auch per Augenschein überzeugen; denn nicht immer gehen hohe P-Werte mit durchgehend niedrigen Restkorrelationen (= Differenzen zwischen beobachteten und vom Modell implizierten Korrelationen, also zwischen S und Σ) einher. Nach WERTS, et al. (1976) sollte daher unterschieden werden zwischen Restkorrelationen, die statistisch bedeutsam sind, und Restkorrelationen, die man für praktisch bedeutsam hält, d.h. beide Beurteilungskriterien (statistische und praktische Signifikanz) sollten angewendet werden.

Zum Test der Hypothesen kann man sich die Additivitätseigenschaft der χ^2-Werte-Verteilungen zunutze machen; denn die Likelihoods der Daten sind stets unter zwei Hypothesen bzw. Modellen miteinander zu vergleichen, von denen das eine Modell in dem anderen enthalten ("nested") ist derart, dass den Parametern des anderen, allgemeineren Modells weitere Restriktionen auferlegt sind. In der Hypothesensequenz bzw. Modellhierarchie

mit zunehmend mehr Restriktionen steigt auch die Anzahl der Freiheitsgrade und der χ^2-Wert. Ein gemessen an der Differenz der Freiheitsgrade zweier Modelle nur wenig veränderter χ^2-Wert zeigt an, dass das restriktivere, also einfachere Modell die Daten mindestens ebenso gut zu reproduzieren vermag. Aus Gründen der Erklärungsökonomie wird dann dem einfacheren Modell unter sonst gleichen Bedingungen der Vorzug gegeben. Die Relationen in der Grösse der Freiheitsgrad- und χ^2-Wert-Differenzen stellen daher einen Test der zusätzlich fixierten Parameter, d.h. des spezielleren Modells dar. Die Differenz in den χ^2-Werten ist eine χ^2-Variable mit Freiheitsgraden gleich der Differenz der Freiheitsgrade der zwei Modelle.

Beim Test der Nullhypothesen muss man davon ausgehen können, dass die jeweiligen Alternativhypothesen die Daten hinreichend genau beschreiben, d.h. die Restkorrelationen müssen durchwegs minimal sein. Denn es ist nicht sinnvoll, die Plausibilitäten zweier Hypothesen miteinander zu vergleichen, die beide an die Daten unangepasst sind. Das gilt auch für den Test der Hypothesen H_4 und H_5 gegen H_3, dem vollständig rekursiven Strukturmodell. In der Pfadanalyse direkt gemessener Variablen (d.h. von "Einzelindikatoren") ist das kein Problem: Dort sind vollständig rekursive Modelle immer an die Daten angepasst. Arbeitet man dagegen mit multiplen Indikatoren, muss man auch bei vollständig rekursiven Modellen mit mangelnder Anpassung rechnen.

Eine Ursache mangelnder Anpassung können im Messmodell nicht berücksichtigte systematische Messfehler sein. Eingangs hatten wir die Vermutung geäussert, aufgrund der besonderen Indikatoren-Zusammenstellung könnte ein Messfehler im Sinne eines Sequenzeffektes (carry-over-effect) entstehen. Dieser Vermutung gehen wir nach, indem wir die Zufallsfehler-Hypothese gegen die allgemeinere Hypothese dieses systematischen Messfehlers testen.

Eine weitere Ursache mangelnder Modellanpassung kann darin liegen, dass die Beziehungen unter den Indikatoren zum grösseren

Teil nicht-linear und/oder nicht-additiv sind. Einheitlich lineare, additive Indikatorenbeziehungen sind notwendige Voraussetzungen dafür, dass latente Variablen im Strukturgleichungs-System konsistent als sozusagen punktuelle Kombinationen ihrer Indikatoren "wirken" können. Darum wird auch die Geltung dieser Voraussetzungen hier überprüft.

Als kritischer P-Wert für die statistische Modellentscheidung wird hier 0,001 angesetzt, d.h. ist einem χ^2-Wert ein $P < 0{,}001$ zugeordnet, wird die Nullhypothese verworfen. In die Festlegung dieses Signifikanzniveaus (die hier vor den Modellvergleichen vorgenommen wird) geht das Wissen darüber ein, dass beim sequentiellen Testen mit folgender Fehlerart zu rechnen ist: Je höher man P ansetzt, desto eher läuft man Gefahr, ein weniger restriktives Modell fälschlicherweise zurückzuweisen.

Der übliche statistische Signifikanzgesichtspunkt wird hier beim sequentiellen Testen nicht allein zur Beurteilung herangezogen. Er wird ergänzt durch einen Gesichtspunkt praktischer Bedeutsamkeit, indem auch die relativen Zuwächse der Restkorrelationsgrössen berücksichtigt werden (s. dazu WERTS, et al., 1976). Gehen in einem konkreten Modellvergleich die Beurteilungen nicht in die gleiche Richtung, geben die relativen Zuwächse der Restkorrelationsgrössen den Ausschlag in der Modellentscheidung. Bei gleichgerichteten Beurteilungen wird hier in der Arbeit nur die statistische explizit erwähnt.

2.2.2. Interne Validitäten und Messfehler-Varianzen der Einzelindikatoren

Die Parameterschätzungen des Messmodells liefern uns die Varianz-Kovarianz-Matrix der latenten Variablen, von der aus die Strukturparameter geschätzt werden können. Bei der Schätzung der Messmodell-Parameter ist danach zu unterscheiden, ob zur Messung der latenten Variablen mehr als ein oder nur ein Indikator (= Einzelindikator) verwendet wird.

Die Messmodell-Parameter der multiplen Indikatoren werden mittels LISREL geschätzt. Ueber diese Parameter ist vor ihrer Schätzung (a priori) nichts bekannt. Den Messmodell-Parametern der Einzelindikatoren dagegen müssen vor der Schätzung der Strukturparameter bestimmte Werte zugewiesen werden. Um die kausalen Zusammenhänge zwischen den latenten Variablen nicht zu unterschätzen, wird hier die Messfehlerbelastetheit des Einzelindikators berücksichtigt.

Die Messfehler-Varianz $Var(\varepsilon_i)$ eines Einzelindikators i lässt sich unmittelbar aus seiner Reliabilität r_{ii} schätzen: $Var(\varepsilon_i) = 1 - r_{ii}$, seine interne Validität λ_i, d.h. seine Korrelation mit der latenten Variablen, ebenfalls: $\lambda_i = \sqrt{r_{ii}}$.

Um diese Werte abzuleiten, könnte auf die von den Testautoren mitgeteilten Koeffizienten der internen Konsistenz, zumeist der Koeffizient α von CRONBACH bzw. Spezialfälle davon, zurückgegriffen werden. Leider fehlen in diesem Untersuchungszusammenhang benötigte Zusatzinformationen. So ist es hier wichtig zu wissen, ob die Komponenten der von uns verwendeten Einzelindikatoren (Items, Subtests oder Testteile) parallel oder zumindest (essential) τ-äquivalent sind. Wenn die Testteile nicht in diesem Sinne homogen sind, dann geben die mitgeteilten Reliabilitätskoeffizienten lediglich eine Schätzung der unteren Grenze für die Reliabilität (LORD & NOVICK, 1968, S. 87 - 90) und mit $\lambda_i = \sqrt{r_{ii}}$ wird dann die interne Validität unterschätzt. Die Quadratwurzel aus dem Koeffizienten α (bzw. aus einem seiner Varianten) ergibt nur dann eine brauchbare Schätzung der internen Validität, wenn mindestens zwei Bedingungen zutreffen:

1. Ein gemeinsamer Faktor erklärt hinreichend die Kovarianzen zwischen den Testteilen,
2. die (unstandardisierten) Regressionen der Testteile auf diesen gemeinsamen Faktor sind gleich bzw. die Kovarianzen zwischen den Testteilen sind gleich.

Setzen sich die verwendeten Einzelindikatoren nicht aus mindestens τ-äquivalenten Komponenten zusammen oder existiert gar mehr als ein gemeinsamer Faktor, dann lässt sich die interne Validität und Reliabilität "verbessern", indem man als Schätzer für die interne Validität den Koeffizienten ρ_{TS} und als Schätzer für die interne Reliabilität den Koeffizienten Ω verwendet. Beide Koeffizienten wurden von HEISE & BOHRNSTEDT (1970) vorgeschlagen. Ω, der Reliabilitätskoeffizient der common-factor-Theorie, und α, der Reliabilitätskoeffizient der true-score-Theorie, sind nur dann gleich, wenn die erwähnten zwei Bedingungen zutreffen (vgl. auch ALLEN, 1974).

Die Messfehler-Varianz eines Einzelindikators wollen wir aber nicht nur adäquat schätzen, sondern durch optimale Linearkombination seiner Teile möglichst auch minimieren. Nur bei Parallelität der Testteile ist die allgemein übliche einfache (= ungewichtete) Summierung der Testteil-Werte gleichzeitig auch die beste, d.h. maximal reliable Linearkombination. Parallele Testteile haben gleiche Varianzen der Messfehler und gleiche Regressionen auf den gemeinsamen Faktor bzw. - um einen in der klassischen Testtheorie (LORD & NOVICK, 1968) gebräuchlichen Ausdruck zu verwenden - gleiche Varianzen der wahren Werte, sind also austauschbar. In diesem Sinne homogene Testteile sind in der Praxis eher selten.

Sind die Testteile nur τ-äquivalent, haben sie also nur gleiche Varianzen der wahren Werte und sind die Varianzen der Messfehler nicht gleich, dann sind sie nicht austauschbar und die unter gegebenen Bedingungen höchstreliable Linearkombination erhält man nur, wenn die Testteile umgekehrt proportional zu ihren Messfehler-Varianzen gewichtet werden.

Ist auch die Annahme gleicher Varianzen der wahren Werte nicht haltbar, sondern nur noch die Annahme eines einzigen

gemeinsamen Faktors ("Ein-Faktor-Hypothese"; JOERESKOG, 1970, bezeichnet in diesem Sinne homogene Testteile als "congeneric", d.h. kongenerische Testteile messen zwar dieselbe Dimension, jedoch nicht in gleichem Masse und nicht mit gleicher Genauigkeit), dann sorgt die Bildung von Faktorwerten für die beste Linearkombination. Die Gewichte der Testteile zur Bildung dieser Linearkombination (Index) sind dann die im Anschluss an eine Faktorenanalyse mit Hilfe der Faktor-Regressions-Methode geschätzten Faktorwert-Koeffizienten. Der Reliabilitätskoeffizient der Wahl (im Rahmen von Analysen der internen Konsistenz) ist dann der Koeffizient Ω.

2.2.3. Kontrollvariable "Geschlecht"

Die Nullhypothesen in den Strukturparametern sollen in jeder Geschlechtergruppe separat getestet werden.

Zum Zwecke dieser Prüfung könnte man mit REVENSTORF (1976, S. 244) folgendes Vorgehen für plausibel halten. Zunächst werden die Parameter (des vollständig rekursiven Strukturmodells) einer Geschlechtergruppe geschätzt. Sodann werden alle geschätzten Parameter fixiert und es wird überprüft, ob das so fixierte Modell auch eine befriedigende Anpassung an die Daten der anderen Geschlechtergruppe liefert.

Dieser Vorschlag REVENSTORFs ist aber nur praktikabel, wenn man bereit ist zu unterstellen, dass die Metrik eines Indikators bzw. einer latenten Variablen in jeder der zu vergleichenden Gruppen dieselbe ist, d.h. es muss vorausgesetzt werden, dass ein Indikator in jeder Gruppe den gleichen Mittelwert

und gleiche Messwertvarianz hat. Doch hat man - bei wachsender Anzahl von Indikatoren um so eher - davon auszugehen, dass die Erfüllung dieser Voraussetzung nicht empirische Regel ist.

Bei der Parameterschätzung mit voller Information ist für die latenten Variablen des Modells eine Metrik festzulegen. Identifizierbar ist ja nur das Verältnis der λ-Parameter (Faktorenladungen) zueinander. Dabei erscheint es plausibel, die Metrik eines Indikators als Standard zu nehmen. Wenn man dementsprechend bei jeder latenten Variablen mit multiplen Indikatoren jeweils einen ihrer λ-Parameter gleich Eins setzt, erhält man unstandardisierte Schätzungen der Strukturparameter (bei LISREL sind das die Maximum-Likelihood-Koeffizienten). Normiert man aber so, dann hängen die Werte der Strukturparameter davon ab, welchen der (multiplen) Indikatoren man als Standard benutzt. Sind aber die Standards, auf die man sich bei der Normierung der latenten Variablen bezieht, in den zu vergleichenden Personengruppen nicht genau gleich, dann sind auch die Werte der Strukturparameter über verschiedene Personengruppen hinweg nicht miteinander vergleichbar. Da uns diese Voraussetzung empirisch kaum gegeben erscheint, wollen wir anders vorgehen. Wir werden die Varianzen der latenten Variablen beim Schätzvorgang auf 1.0 fixieren und die Messmodell-Parameter der multiplen Indikatoren nur auf gleiche, aber unbekannte Werte (Gleichheitsrestriktionen unter der Annahme paralleler Indikatoren). Die mit der Korrelationsmatrix geschätzten Strukturparameter sind dann standardisierte Effektparameter oder Pfadkoeffizienten. Analoges gilt auch für die Messmodell-Parameter. Diese sind dann standardisierte Faktorenladungen. Dies stimmt mit der Interpretation der Messmodell-Parameter bei Einzelindikatoren überein: Die interne Validität eines Einzelindikators, hier geschätzt als Quadratwurzel aus seiner Reliabilität, ist als standardisierte Faktorenladung aufzufassen (standardisierte Regression dieses Indikators auf

die latente Variable bzw. Korrelation dieses Indikators mit der latenten Variablen).

Zum Vergleich von Kausalmodellen über verschiedene Personengruppen hinweg wird von einigen Autoren (u.a. TUKEY, 1954; BLALOCK, 1968 b; SCHOENBERG, 1972) empfohlen, mit unstandardisierten Effektparametern zu arbeiten. Doch für unsere Zwekke des Vergleichs erscheint uns eher die Verwendung standardisierter Parameter vertretbar, und zwar aus folgenden Gründen:

1. Direkte Vergleiche der absoluten Schätzwerte der Strukturparameter streben wir nicht an.

   Wir wollen den Kausalmodellen nicht die Information entnehmen, um wieviel die Veränderung einer Variablen um jeweils eine Standardabweichung ceteris paribus eine davon abhängige Variable verändern würde. In Anbetracht der Fehlerbelastetheit sozialwissenschaftlicher Messung ist das Forschungsziel hier bescheidener: Hier genügt es nur zu vergleichen, ob bestimmte a priori postulierte Effekte vorhanden sind oder nicht, mit anderen Worten: ob die Nullrestriktionen, die den Strukturparametern bei der ersten Gruppe ohne Anpassungsverschlechterung auferlegt wurden, von der anderen Gruppe auch in diesem Sinne übernommen werden können. Denn standardisierte und nicht-standardisierte Effektparameter stimmen überein, sobald einer von ihnen Null wird (Formel z.B. bei BLALOCK, 1968 b, S. 190). Dies gilt zwar nicht theoretisch, praktisch aber immer. Das bedeutet, dass wir Modellvergleiche dieser Art im Rahmen von LISREL auch mit der Korrelationsmatrix der Indikatoren (statt mit der Varianz-Kovarianz-Matrix) durchführen können, d.h. die statistische Beurteilung der Modellvergleiche bzw. der χ^2-Wert-Differenzen ist hier von der Standardisierung der Variablen unabhängig.

2. Ebenso genügt es, nur die Gewichtsverhältnisse der Messmodell-Parameter bzw. Faktorenladungen zu kennen, ohne dass es auf die genaue Höhe der Gewichte selbst ankommt.

   Der soeben vorgeschlagene Gruppenvergleich von Strukturparametern ist nur dann zulässig, wenn die Struktur des Messmodells in beiden Geschlechtergruppen dieselbe ist, d.h. die Bedeutung einer latenten Variablen darf sich über beide Geschlechtergruppen hinweg nicht wesentlich verändern.

   Bezogen auf die Einzelindikatoren heisst dies, dass die interne Validität eines Indikators in beiden Geschlechtergruppen im wesentlichen gleich sein muss. Davon hat man sich vor dem Gruppenvergleich von Strukturparametern zu überzeugen. Für die multiplen Indikatoren gelten etwas andere Bedingungen, da deren Parameter mit den Strukturparametern simultan geschätzt werden. Im allgemeinen faktorenanalytischen Modell erhalten die Faktoren bzw. latenten Variablen ihre empirische Bedeutung durch diejenigen Indikatoren, die hoch auf ihnen laden. Das Gewichtsverhältnis der Faktorenladungen zueinander bestimmt dann, an welchen Indikatoren im einzelnen sich diese Bedeutungsgebung zu orientieren hat. In dieser Hinsicht, das hatten wir eingangs zur Verwendbarkeit multipler Indikatoren festgelegt, sollten die derselben latenten Variablen zugeordneten Indikatoren einander äquivalent sein (Bedingung paralleler Indikatoren, s. S. 62). Gerade diese Aequivalenzbedingung, separat in jeder Geschlechtergruppe überprüft, gewährleistet auch Bedeutungs- oder Interpretations-Invarianz hinreichend. Entscheidend ist nur die numerische Uebereinstimmung der Faktorenladungen von Indikatoren derselben latenten Variablen innerhalb jeder Geschlechtergruppe. Auf die Uebereinstimmung in den absoluten Werten entsprechender Faktorenladungen zwischen den Geschlechtergruppen

kommt es dabei nicht notwendig an. Sind die Gewichtsverhältnisse einander zugeordneter Indikatoren über verschiedene Personengruppen hinweg konstant, wollen wir von "struktureller Homogenität" der Messmodelle sprechen.

2.2.4. Messniveau und Analyseverfahren

Die Frage ist, ob die Messverfahren den Voraussetzungen genügen, die wir der statistischen Verarbeitung der Daten zugrundelegen. Wenn wir nicht nur die Richtung, sondern auch die Form der Beziehungen zwischen den latenten Variablen spezifizieren, dann impliziert eine solche Präzisierung der Hypothesen metrische Eigenschaften der latenten Variablen und mindestens Intervallskalenqualität der Indikatoren als den zwar unbekannten, aber linearen Transformationen dieser latenten Variablen.

Auch wenn wir mit LISREL ein Verfahren anwenden, das intervallskalierte Indikatoren voraussetzt, so wollen wir doch unsere Daten nicht für seriöser ausgeben als sie sind: Es besteht Ungewissheit darüber, ob alle verwendeten Masse auf einer Intervallskala liegen. Da bei einigen Messverfahren die interne Kontrolle (bzw. empirische Prüfung bestimmter Repräsentationstheoreme, s. dazu SUPPES & ZINNES, 1963) der von den Autoren offensichtlich angezielten Intervallmessung unterblieb, wird von der Struktur der latenten Variablen vermutlich häufig nicht mehr erfasst als nur die Ordnung.

Dennoch sehen wir keinen entscheidenden Grund, uns bei Daten mit möglicherweise mässigen Skalenqualitäten auch mit mässiger Analyse zu begnügen. Zum einen gibt es zum LISREL-Verfahren keine gleichwertige ordinale Alternative. Sie gibt es noch nicht einmal, wie KIM (1975) anhand einfachster Beispiele gezeigt hat, für Kausalmodelle mit direkt gemessenen Variablen.

Zum anderen gehen wir davon aus, dass für alle Messoperationen immer auch ihre empirische Rechtfertigung durch Verknüpfung mit Aussenkriterien (externe Kontrolle der Messung) gegeben werden muss. Dies kann im LISREL-Ansatz geschehen, wenn dieselbe Variable mehrfach gemessen wird: Reproduziert das postulierte Kausalmodell die Daten, d.h. führt die Verwendung alternativer Messoperationen jedesmal zu gleichen Ergebnissen der Hypothesenprüfung, dann werden wir diese Ergebnisse nicht deswegen als nicht aussagekräftig ansehen, weil die Transformationseigenschaften der Messverfahren unbekannt sind. Die Behandlung unserer Messverfahren als Intervallskalen bzw. die Angemessenheit von LISREL wird durch den Erfolg, durch die sinnvolle und richtige Deutbarkeit der Ergebnisse gewissermassen im nachhinein bestätigt. Wir halten das Uebersehen eines praktisch bedeutsamen Ergebnisses für folgenschwerer als uns aus Gründen der messtechnischen Vorsicht im vorhinein gegen die Anwendung von LISREL zu entscheiden. Wir meinen, dass es unter diesen Gesichtspunkten zunächst einmal gleichgültig ist, nicht genau zu wissen, ob die Daten dem von LISREL geforderten Messniveau überhaupt genügen oder nicht.

2.3. Durchführung

2.3.1. Ablauf der Untersuchung

Die Untersuchung fand 1975 in einer Gesamtschule der Stadt Wetzlar statt. Erfasst wurde der gesamte 6. Schuljahrgang am Ende seiner Förderstufenzeit.

Im Kreise Wetzlar ist die Förderstufe seit dem 1. 8. 1969 obligatorisch, d.h. man richtete an allen Hauptschulen, Realschulen und Gymnasien des Kreises Förderstufen ein, die alle Schüler der Schuljahrgänge 5 und 6 zu besuchen hatten.

Der hier gestartete Schulversuch erprobt die Förderstufe (bzw. die Gesamtschule) im Flächenversuch, also unter Regelschulbedingungen.

Bei der Auswahl der Schule spielten folgende Gesichtspunkte eine Rolle (s. auch S. 32):

1. Der Radius des Einzugsbereichs der Schule sollte die ganze Stadt Wetzlar einschliessen.
2. Die Schülerpopulation sollte in Bezug auf die soziale Herkunft ungefähr den Grössenverhältnissen in der BRD entsprechen.

Die Vorinformationen dazu wurden bei der Schulaufsichtsbehörde in Wetzlar eingeholt und dementsprechend wurde die Schule ausgewählt.

Alle Eltern der von uns befragten Schüler wurden in einem Schreiben darauf hingewiesen, dass die Teilnahme ihrer Kinder an unserer Befragung freiwillig ist. Das Einverständnis der betroffenen Stellen (hessisches Kultusministerium, Schulaufsichtsbehörde in Wetzlar) und Personen (Lehrer und Eltern der befragten Schüler) zur Befragung erfolgte nur unter der Voraussetzung, dass die Anonymität der Schüler abgesichert werden konnte. Die anonyme Zuordnung getrennt anfallender Daten geschah über vierstellige Schülerkennziffern. Das "Deutsche Institut für Internationale Pädagogische Forschung" (DIPF) in Frankfurt/M. besass eine Liste mit Schülernamen und Kennziffern. Vor Beginn der Untersuchung wurde jedem Schüler seine Kennziffer mitgeteilt (durch einen Lehrer, ohne Beisein der Versuchsleiter). Auf jedem der zu bearbeitenden Fragebögen war diese Ziffer zu notieren. Auf diese Weise konnten die später vom DIPF übermittelten Daten zur sozialen Herkunft des Schülers mit seinen Fragebogendaten in Beziehung gesetzt werden, ohne dass eine namentliche Identifizierung notwendig war.

In der ausgewählten Schule war es aus technischen Gründen nicht möglich, alle Schüler des 6. Schuljahrganges innerhalb eines einzigen Raumes zu befragen. Darum wurden die Schüler auf 6 Klassenräume verteilt. Die 6 Gruppen wurden nach den Kurszugehörigkeiten der Schüler zusammengestellt: zwei A-Kurs-, zwei B-Kurs- und zwei C-Kurs-Gruppen.

Damit die Befragung der Gruppen zur gleichen Zeit stattfinden konnte, wurden 5 weitere Versuchsleiter engagiert (für eine Schülergruppe war der Verfasser dieser Arbeit selbst Versuchsleiter). Es waren dies 5 Diplom-Psychologen vom Psychologischen Institut der Universität Giessen. Lehrer der betreffenden Schule sollten mit der Befragung nicht betraut werden und Lehrer durften während der Befragung auch nicht im Raume anwesend sein. Diese Untersuchungsbedingung erschien uns psychologisch wichtig: Wir gingen davon aus, dass von einer deutlichen Demonstration der Anonymität des einzelnen Schülers ganz entscheidend die Aussagewilligkeit und Aussagefähigkeit der Schüler abhängt, insbesondere wenn es um Aussagen zum Eltern- und Lehrerverhalten geht.

Vor der Untersuchung fanden mehrere Sitzungen statt, in denen die Versuchsleiter auf ihre Aufgaben vorbereitet wurden. Sie wurden mit den Messverfahren (bzw. ihren Instruktionen) vertraut gemacht, schriftlich fixierte Arbeitsanweisungen wurden ausgehändigt und besprochen. Die Versuchsleiter wurden gebeten, sich streng an diese Anweisungen zu halten, damit alle 6 Untersuchungsgruppen unter gleichen Durchführungsbedingungen arbeiten konnten.

Die wichtigsten Arbeitsanweisungen waren:

1. Der Versuchsleiter achtet darauf, dass die Schüler gewissenhaft und sorgfältig mitarbeiten. Wenn ein Schüler Widerstände zeigt mitzuarbeiten, dann wollen wir das so regeln: keiner der Schüler darf den Klassenraum während der Befragung unerlaubt verlassen (zum Verständnis des

Wortes "unerlaubt" in Anbetracht der Freiwilligkeit der Teilnahme sei hier eingefügt: Hätte die Befragung nicht stattgefunden, wäre normaler Schulunterricht gewesen. Drei reguläre Schulstunden, genauer: die von 8.00 bis 10.40 Uhr, fielen für diese Befragung aus). Es wird diesem Schüler jedoch freigestellt, ob er in der noch laufenden Schulstunde den Fragebogen ausfüllen oder aber sich mit anderen Arbeiten still beschäftigen will.

2. Die Instruktionen zu den einzelnen Verfahren sind möglichst wörtlich zu geben. Es ist unter allen Umständen darauf zu achten, dass die Schüler die Anweisungen richtig verstehen. Wenn einige Schüler ratlos dasitzen, muss sie der Versuchsleiter direkt ansprechen und nach ihren Schwierigkeiten fragen. Nach Beendigung der Instruktionen bzw. Uebungsbeispiele fragt der Versuchsleiter jedesmal, ob jeder Schüler weiss, was er zu tun hat.

3. In den Schulpausen und am Ende der Befragung schaut der Versuchsleiter die Fragebögen im Ueberblick durch um sich zu vergewissern, ob wirklich alle Fragen beantwortet wurden. Bemerkt der Versuchsleiter ausgelassene Sätze oder falsche Antworten (z.B. zwei Antworten für eine Frage) oder fehlende Schülerkennziffern, so bittet er um entsprechende Korrektur.

4. Feste Reihenfolge der Aufgaben in allen 6 Gruppen:
   1. Allgemeine Einführung
   2. Personalstatistische Daten
   3. Mutter-Unterstützungsskala
   4. Vater-Unterstützungsskala
   5. Anstrengungsvermeidungstest
   6. Lehrer-Unterstützungsskala
   7. Lehrer-Sympathieskala
   8. Intelligenztest

   Diese Abfolge ist zeitlich so eingerichtet, dass die re-

gulären (kurzen) Schulpausen eingehalten werden können. Der Versuchsleiter sollte darauf achten, dass das Pausenzeichen der Schule nicht mitten in die Durchführung eines Tests fällt, d.h. vor dem Ende der Schulstunde sollte kein neuer Test begonnen werden, wenn die Aussicht besteht, dass er in die Pause hinein durchgeführt werden muss.

2.3.2. Versuchspersonen

Der untersuchte Schuljahrgang bestand aus insgesamt 162 Schülern. 18 davon gelangten nicht in die Auswertung, und zwar aus folgenden Gründen:

- Fernbleiben krankheitshalber
- unvollständige Angaben
- zu grosse Sprachschwierigkeiten (Ausländerkinder)

Bei Ausfällen aus diesen Gründen sind die in die Auswertung übernommenen 144 Schüler durchaus als Zufallsauswahl aus der grösseren Gesamtheit "Schuljahrgang am Ende der Förderstufe" aufzufassen.

In dieser Auswahl befanden sich 74 Jungen und 70 Mädchen. Das Alter dieser Schüler reichte von minimal 11;3 bis maximal 14;9 Jahren. Der Altersdurchschnitt lag bei 12;5 Jahren.

Die soziale Herkunft der 144 Schüler, klassifiziert nach dem BRSS von BAUER und reduziert auf drei Stufen, verteilte sich wie folgt:

I : Oberschicht und obere Mittelschicht: 27 (= 19%)
II : mittlere und untere Mittelschicht : 48 (= 33%)
III: obere und untere Unterschicht : 69 (= 48%)

Vergleicht man unsere Zahlen mit den nach dem SCHEUCH-Index für die BRD ermittelten Zahlen - BRSS von BAUER und SCHEUCH-Index sind hinsichtlich der Anzahl und der Bezeichnungen der Schichten direkt miteinander vergleichbar -, dann zeigt sich, berücksichtigt man den kleineren Umfang unserer Stichprobe, eine annähernde Uebereinstimmung der Verteilungen: Stufe I: 13%, Stufe II: 42% und Stufe III: 45% (DAHRENDORF, 1968, S. 99).

Das bei unserer Stichprobe vergleichsweise heterogenere Spektrum sozialer Herkunft sollte man nicht überbewerten. Zahlenvergleiche dieser Art sind nur mit Vorbehalten aufzunehmen, u.a. weil mit verschiedenen Schichtansätzen nicht direkt vergleichbare Interpretationsschemata mitgegeben sind, die die Grenzziehung zwischen den einzelnen Schichten und damit deren Umfänge verschieben. So könnte die höhere Belegung von Stufe I (19% versus 13%) ceteris paribus damit zu erklären sein, dass der als Vehikel zum Berufsaufstieg heute häufiger genutzte Indikator "Schulbildung" im BRSS von BAUER stärker gewichtet wird als im SCHEUCH-Index. Aus dem Vergleich jedoch lässt sich zumindest ableiten, dass im untersuchten Schuljahrgang keine der Stufen sozialer Herkunft stark unter- bzw. über-repräsentiert ist.

Die Verteilung der Geschlechter auf die drei Kursniveaus A, B und C in den Fächern Mathematik und Sprachen zum Zeitpunkt der Untersuchung zeigt Tabelle 1. Bei dieser Darstellung wurden nur die in die Auswertung gekommenen 144 Schüler berücksichtigt. Bei den übrigen 18 Schülern gehen wir davon aus, dass sie sich unsystematisch auf die Kursniveaus verteilen.

16 von den 144 Schülern hatten während der gesamten Förderstufenzeit anstelle von Englisch einen Latein-Kurs belegt. Nach Auskunft der von uns befragten Lehrer ist die Teilnahme

am Latein-Unterricht der A-Kurs-Zugehörigkeit in Englisch äquivalent. Zur Vereinfachung der folgenden Auswertungen fassen wir daher beide Möglichkeiten in der Gruppe "A-Kurs-Zugehörigkeit in Sprachen" zusammen. B- bzw. C-Kurs-Zugehörigkeit in "Englisch" ändern wir entsprechend um in: B- bzw. C-Kurs-Zugehörigkeit in "Sprachen".

Tab. 1: Verteilung der Geschlechter auf die drei Kursniveaus A, B und C (A = höchstes, C = niedrigstes Kursniveau) in Mathematik und Sprachen. Zeitpunkt: Ende der Förderstufe (N = 144 Schüler)

	Mathematik			Sprachen		
Geschlecht	A	B	C	A	B	C
Jungen	23	28	23	24	27	23
Mädchen	34	19	17	35	24	11
Gesamt	57	47	40	59	51	34

Jungen und Mädchen verteilen sich augenscheinlich nicht gleichmässig auf die verschiedenen Kursniveaus: Unabhängig vom Fach sind Mädchen häufiger im höchsten Kursniveau, Jungen dagegen häufiger im niedrigsten Kursniveau vertreten. Der letzte Sachverhalt hat auch im Sonderschulbereich seine Parallele: Im Verhältnis von ca. 3 : 2 besuchen Jungen gegenüber Mädchen die Sonderschule, obwohl nicht davon ausgegangen werden kann, dass die Testintelligenz dieser Jungen niedriger ist als die der Mädchen (MUELLER, 1970).

Unseren numerischen Eindruck müssen wir nicht unbedingt statistisch bestätigen: Die Ueber- bzw. Unter-Repräsentation

einer Geschlechtergruppe in bestimmten Kursniveaus führt für den Test der Nullhypothesen nicht zu Fehlurteilen hinsichtlich geschlechtsbedingter Scheinkorrelationen, da von Anfang an nach Geschlechtergruppen getrennt getestet wird.

Die Verteilung aller Schüler auf die drei Kursniveaus in Mathematik weicht nur unwesentlich von derjenigen in Sprachen ab. Bei einer solchen Verteilung liegt die Vermutung nahe, dass die meisten der untersuchten Schüler in beiden Fächern dem gleichen Kursniveau angehören ("Bankbildung"). Als Gradmesser der Bankbildung kann man die Höhe der Korrelation zwischen den Kurseinstufungen in Mathematik und Sprachen ansehen. Die Produkt-Moment-Korrelation zwischen diesen (als gleichabständig gedachten) Kurseinstufungen beträgt bei den Jungen $r = 0,73$ und $r = 0,82$ bei den Mädchen. Diese Korrelationen stützen die Vermutung einer Bankbildung (ein statistischer Signifikanztest erübrigt sich hier wohl) und offensichtlich ist diese Bankbildung vom Geschlecht der Schüler unabhängig.

Betrachtet man nur die Gruppe derjenigen Schüler, die in mindestens einem Leistungsfach dem niedrigsten oder höchsten Kursniveau angehören, so verstärkt sich aus dieser Perspektive der Eindruck einer Bankbildung noch. Es zeigt sich: Wer schon in einem Fach dem A-Kurs angehört, der gehört mit absoluter Wahrscheinlichkeit nicht dem C-Kurs des anderen Faches an (sogenannte "obere" Leistungsgruppe). Wer in schon einem Fach dem C-Kurs angehört, der gehört ebenso mit Sicherheit nicht dem A-Kurs des anderen Faches an (sogenannte "untere" Leistungsgruppe). Man kann also davon sprechen, dass am Ende der Förderstufenzeit ein relativ undifferenziertes, fachübergreifendes, d.h. quasi-streaming organisiertes Kurssystem besteht.

Die Berücksichtigung nur dieses Zeitpunktes im Rahmen des vorgegebenen Kausalmodells setzt voraus, dass eben dieses Quasi-streaming über die Förderstufenzeit hinweg relativ stabil war. Vergleichen wir nun die Ersteinstufung mit der am Ende der För-

derstufenzeit bestehenden Einstufung, getrennt für Schüler der unteren und oberen Leistungsgruppe, so ist festzustellen: Schüler, die am Anfang der Förderstufe der unteren Leistungsgruppe angehören (n = 40), gehören ihr ohne Ausnahme auch am Ende der Förderstufe an. Es findet kein "Aufstieg" in die mittlere (B/B-Kurszugehörigkeit) oder obere Leistungsgruppe (also in mindestens einem Fach im A-Kurs) statt. Eine ähnliche Umstufungspraxis, nur mit umgekehrtem Vorzeichen, lässt sich bei der oberen Leistungsgruppe beobachten: Von den Schülern, die am Anfang der Förderstufe der oberen Leistungsgruppe angehören (n = 75), wird am Ende der Förderstufe keiner in den C-Kurs herabgestuft, auch nicht in einem Fach. Ein so praktiziertes Kurssystem ist also in relativ stabiler Weise quasi-streaming organisiert.

Zusammenfassend: Für unser Messmodell erscheint es deshalb nicht nur gerechtfertigt, dass die Kurszugehörigkeiten am Ende der Förderstufe auf einen gemeinsamen Faktor (schulische Leistungsfähigkeit) zurückgeführt werden. Man kann diesen Faktor darüber hinaus auch als stellvertretend für "die" Kurszugehörigkeit während der Förderstufenzeit ansehen. Das Vorliegen relativ stabiler quasi-streaming organisierter Kurszugehörigkeiten ist Voraussetzung der beabsichtigten kausalanalytischen Hypothesenüberprüfungen mit LISREL.

3. ERGEBNISSE

3.1. Voruntersuchungen

3.1.1. Index-Bildung

Dem Verfahren von BRUSTEN & HURRELMANN zur Erfassung der Lehrerunterstützung liegt kein ausgearbeitetes Mess- oder Skalierungsmodell zugrunde. Ausserdem werden von den Autoren kaum teststatistische Angaben gemacht. Aus der eher vagen Angabe, dass die Items des Verfahrens hoch miteinander korrelieren, lässt sich zwar ableiten, dass die Items womöglich ein- und dieselbe Eigenschaft messen, jedoch nicht, dass die übliche einfache Addition der Itemwerte zum Gesamtpunktwert gleichzeitig auch die beste, d.h. maximal reliable oder valide Linearkombination ist. Das ist nur bei Parallelität der Items der Fall. Es ist darum notwendig, die Items mit Hilfe unserer Daten unter den verschiedenen Homogenitätsbedingungen (s. S. 79) zu analysieren.

Ziel dieser Analyse ist primär, eine innerhalb des Strukturmodells brauchbare Linearkombination der Items (Index) zu bilden. Im Strukturmodell stellt dieser Index einen multiplen Indikator dar. Da sich die Qualität multipler Indikatoren, ablesbar an den Faktorenladungen bzw. den Messfehler-Varianzen, vor allem aus der Anwendung von LISREL simultan mit der Ueberprüfung der eigentlichen Hypothesen ergibt, kann eine dieser Hypothesenprüfung vorangehende, nach Geschlechtern getrennte Qualitätskontrolle der Einfachheit halber entfallen. Mit anderen Worten, die im Anschluss an diese Index-Bildung geschätzte interne Reliabilität bzw. Validität taucht im Gegensatz zu den Einzelindikatoren nicht als explizit fixierte Grösse im Modell multipler Indikatoren auf.

Die Varianz-Kovarianz-Matrix der 5 Items der Lehrer-Sympathie- (LESY) bzw. Lehrer-Unterstützungs-Skala zeigt Tabelle 2:

Tab. 2: Varianz-Kovarianz-Matrix der 5 Items der Lehrer-Sympathie-Skala (LESY) von BRUSTEN und HURRELMANN (N = 144)

Items	1	2	3	4	5
1	1,356				
2	0,563	1,360			
3	0,406	0,490	1,619		
4	0,785	0,689	0,228	1,499	
5	0,391	0,603	0,346	0,476	1,103

Die LISREL-Analyse dieser Daten unter den drei Homogenitätsbedingungen: parallel, τ-äquivalent und kongenerisch, ergab die in der folgenden Tabelle angeführten Werte:

Tab. 3: Analyse der Varianz-Kovarianz-Matrix von Tab. 2 unter drei verschiedenen Homogenitätsbedingungen

Hypothesen	df	χ^2	P
parallel	13	42,7143	0,0001
τ-äquivalent	9	26,2861	0,0018
kongenerisch	5	16,5900	0,0053

Nur unter der Parallelitäts-Bedingung wird der χ^2-Wert signifikant ($P < 0{,}001$). Daraus leiten wir ab, dass die Items zwar nicht parallel, doch zumindest homogen im kongenerischen bzw. einfaktoriellen Sinne sind.

Es sieht so aus, als sei die Anpassung des Modells τ-äquivalenter Items, gemessen am allgemeineren Ein-Faktor-Modell, nur geringfügig schlechter. Auch statistisch wird dieser Eindruck bestätigt: Die Differenz der χ^2-Werte: 26,2861 - 16,59 = 9,6961 liegt bei 9 - 5 = 4 Freiheitsgraden zwar ziemlich knapp unter der Signifikanzschranke ($0{,}001 < P < 0{,}05$), doch kann die Hypothese einer Anpassungsverschlechterung durch das Modell τ-äquivalenter Messungen noch zurückgewiesen werden. Dennoch wollen wir das einfachere Messmodell preisgeben: Wir halten nämlich die Anpassungsverbesserung, die bei Lockerung seiner Gleichheitsrestriktionen eintritt, für praktisch bedeutsam.

Dies lässt sich einmal durch den Vergleich der (hier nicht dargestellten) Residualmatrizen ($\Sigma - S$) beider Messmodelle belegen. Da Residualkorrelationen leichter zu interpretieren sind als Residuen von Varianzen und Kovarianzen, werden wir im folgenden von den Residualkorrelationen ausgehen. Als (deskriptiven) Kennwert für das Ausmass der Abweichungen der beobachteten von den vom Messmodell implizierten Korrelationen kann man die Quadratwurzel aus dem Durchschnitt der quadrierten Differenzen (der Korrelationen) ansehen. Danach zeigt sich, dass unter der restriktiveren τ-Aequivalenz-Bedingung die Residualkorrelationen durchschnittlich doppelt so hoch liegen (0,12) wie unter der allgemeinen Ein-Faktor-Bedingung (0,06). Die grösste absolute Differenz ist unter der τ-Aequivalenz-Hypothese: 0,239 und unter der Ein-Faktor-Hypothese: 0,115. Diese Unterschiede halten wir für praktisch bedeutsam. Wir vertrauen also weniger der τ-Aequivalenz-Hypothese, obwohl das Ergebnis des statis-

tischen Signifikanztests knapp in diese Richtung weist, und akzeptieren die Hypothese bloss kongenerischer Items.

Die praktische Bedeutsamkeit der eben festgestellten Unterschiede zeigt sich auch, wenn wir die unter der Hypothese kongenerischer Items geschätzten Faktorenladungen β_i in Tabelle 4 betrachten:

Tab. 4: Maximum-Likelihood-Schätzungen der Faktorenladungen β_i und ihrer Standardschätzfehler σ_i unter der Hypothese kongenerischer Items der Lehrer-Sympathieskala

Item i	β_i	σ_i
1	0,761	0,099
2	0,840	0,098
3	0,479	0,117
4	0,850	0,104
5	0,615	0,091

Man sieht, dass die Faktorenladungen β_i auch unter Berücksichtigung ihrer Standardschätzfehler σ_i voneinander abweichen. Es ist augenscheinlich, dass die Abweichung von der Annahme gleicher Faktorenladungen vor allem zu Lasten von Item 3 geht. Wiederum erscheint also, da wir Item 3 nicht eliminieren wollen, die τ-Aequivalenz-Bedingung weniger plausibel.

Offensichtlich messen die Items dieselbe Eigenschaft nicht nur unterschiedlich genau, sondern auch in unterschiedlichem Masse. Unter diesen Umständen erscheint eine Index-Bildung angezeigt, welche die Items nicht nur entsprechend ihren Messfehler-Varianzen (wie es bei τ-äquivalenten Items

angebracht wäre), sondern auch entsprechend ihren Faktorenladungen gewichtet.

Die zur optimalen Summenbildung benötigten Itemgewichte sind die mit Hilfe der Faktor-Regressions-Methode geschätzten Faktorwert-Koeffizienten. Nun unterscheiden sich die so gebildeten Linearkombinationen danach, welches faktorenanalytische Modell wir ihrer Schätzung zugrundelegen. Wegen der teilweisen Inkompatibilität von Reliabilität und Validität gibt es zudem keine Faktor-Regressions-Methode, die die Faktorwert-Koeffizienten so bestimmt, dass (interne) Reliabilität und Validität des zu konstruierenden Faktorwertes (Index) simultan maximiert werden. Mit der Wahl der Methode wird also auch eine Entscheidung darüber verlangt, was maximiert werden soll: die Reliabilität oder die Validität.

Da es uns bei multiplen Indikatoren auch darauf ankommen muss, ihre Brauchbarkeit innerhalb des Strukturmodells zu erhöhen, ist hier vor allem das Gültigkeitsproblem angesprochen. Ist bei der Index-Bildung die Betonung auf Maximierung der Validität zu legen, dann ist die kanonische Faktorenanalyse von RAO (1955) das Verfahren der Wahl: Die kanonische Faktor-Regressions-Methode ist ein die Faktorenwerte eindeutig schätzendes Verfahren, da es die Korrelation zwischen den konstruierten Faktorwerten mit den ihnen entsprechenden wahren Werten maximiert (den Beweis führt: McDONALD, 1968).

Die mit der kanonischen Faktorenanalyse von RAO geschätzten Faktorwert-Koeffizienten w_i und Faktorenladungen a_i (es wurde im SPSS-Programmsystem von NIE et al., 1975, gerechnet) sind in Tabelle 5 aufgeführt:

Tab. 5: Kanonische Faktorwert-Koeffizienten w_i und Faktorenladungen a_i der 5 Items der Lehrer-Sympathieskala, ferner ihre arithmetischen Mittel $\bar{x}_i$ und Standardabweichungen s_i.

Item i	w_i	a_i	$\bar{x}_i$	s_i
1	0,257	0,637	2,813	1,165
2	0,303	0,690	3,868	1,166
3	0,119	0,386	3,743	1,272
4	0,279	0,664	3,340	1,224
5	0,217	0,582	3,958	1,050

Die gesuchte Linearkombination zur Bildung der Faktorwerte - wir wollen sie abgekürzt auch hier "LESY" nennen, d.h. ab jetzt steht "LESY" (s. S. 64/65) nicht mehr wie bisher für die einfache, sondern für die gewogene Addition der Itemwerte - sei der Deutlichkeit halber explizit angeschrieben:

$$\begin{aligned} LESY = {} & 0{,}257 * (x_1 - 2{,}813)/1{,}165 + \\ & 0{,}303 * (x_2 - 3{,}868)/1{,}166 + \\ & 0{,}119 * (x_3 - 3{,}743)/1{,}272 + \\ & 0{,}279 * (x_4 - 3{,}340)/1{,}224 + \\ & 0{,}217 * (x_5 - 3{,}958)/1{,}050 \end{aligned}$$

wobei: x_i = Rohwert von Item i, i = 1,2,3,4,5

Interne Reliabilität Ω und interne Validität ρ_{TS} der obigen Linearkombination (Index) sind in diesem Zusammenhang, d.h. nur wenn mit der kanonischen Faktorenanalyse gerechnet wird, sehr einfach zu schätzen:

$$\Omega = \Sigma(a_i * w_i) = 0{,}730$$

$$\rho_{T\cdot} = \sqrt{\Omega} = 0{,}854$$

Die Ableitung dieser Beziehungen findet sich bei ALLEN (1974).

3.1.2. Interne Validitäten und Messfehler-Varianzen der Einzelindikatoren

3.1.2.1. Intelligenztest (CFT2)

Der von WEISS (1972) zur Reliabilitätsschätzung des CFT2 errechnete Koeffizient der internen Konsistenz (Testteile hier: die 4 Subtests) wurde an einer Gruppe 11jähriger Schüler des 5. Schuljahrganges gewonnen. Unabhängig von der ungeklärten Frage, ob die geforderten Homogenitätsbedingungen (mindestens essentielle τ-Aequivalenz der Subtests) überhaupt vorliegen: Es ist ungewiss, ob diese Reliabilitätsschätzung auf andere Altersgruppen verallgemeinerbar ist, also auch für 12 - 14jährige Schüler gilt, mit denen wir es in unserer Untersuchung hauptsächlich zu tun haben. Je nach Referenzpopulation sind verschiedene, d.h. populationsabhängige Beurteilungen der Zuverlässigkeit zu erwarten. Die vor der Hypothesenprüfung festzulegenden Messmodell-Parameter des CFT2 (interne Validität λ_1 und Messfehler-Varianz $Var(\varepsilon_1)$) werden wir deshalb aus den eigenen Daten ableiten.

Auch kann die Reliabilität des CFT2 vom Geschlecht der Schüler abhängen. Dies müssten wir berücksichtigen, wenn wir die Geschlechtsabhängigkeit der Strukturparameter untersuchen. Wir müssen also wissen, ob die zur Reliabilitätsschätzung (und das heisst auch: die zur Schätzung der internen Validität und Messfehler-Varianz) benötigten Homogenitätsvoraussetzungen (Parallelität und/oder τ-Aequivalenz) in jeder unserer Geschlechtergruppen gleichermassen gelten.

Die Homogenität der Homogenitätsvoraussetzungen überprüften wir mit dem Computerprogramm COFAMM von JOERESKOG & SOERBOM (1976). Mit COFAMM kann eine Hypothese in mehreren

Populationen gleichzeitig beurteilt werden.

Zunächst testeten wir die Homogenität der Parallelitäts-Bedingung. Ausgehend vom Ein-Faktor-Modell: $\Sigma = \beta\beta' + \Theta^2$ fixierten wir die β- und Θ^2-Parameter der $m = 4$ Subtests des CFT2 auf jeweils gleiche, aber unbekannte Werte, so dass insgesamt nur zwei freie Parameter zu schätzen waren. Unter diesen Restriktionen wurden die Varianzen und Kovarianzen der Subtests analysiert, welche für Jungen und Mädchen getrennt in Tabelle 6 aufgeführt sind.

Die Rohwerte der Subtests, nicht deren IQ-Werte, bildeten die Grundlage für die Berechnung der Varianzen und Kovarianzen. Auch in unseren weiteren Analysen wurde auf die von WEISS angebotene Altersnormierung der CFT2-Werte verzichtet.

Tab. 6: Varianz-Kovarianz-Matrix der 4 Subtests des CFT2 (1. Testteil) von WEISS, aufgegliedert nach Geschlecht

	Subtests	1	2	3	4
Jungen	1	3,093			
	2	1,980	3,858		
	3	1,747	1,960	3,790	
	4	1,335	1,670	0,766	3,632
Mädchen	1	3,757			
	2	2,061	4,875		
	3	2,554	2,253	4,576	
	4	1,635	2,579	2,074	3,686

Der statistische Test der Hypothese der Parallelitäts-Invarianz ergab einen Wert von $\chi^2 = 24{,}538$, der bei 17 Freiheitsgraden einem $P = 0{,}1055$ entspricht. Die Hypothese, dass der CFT2 (hier der 1. Testteil) aus lauter parallelen Komponenten besteht und dass diese Parallelität darüber hinaus in beiden Geschlechtsgruppen gleich ausgeprägt ist, lässt sich also statistisch nicht widerlegen.

Daraus leiten wir ab, dass zur Reliabilitätsschätzung Varianz- und Kovarianz-Unterschiede über beide Geschlechtergruppen hinweg vernachlässigbar sind. Die beste Linearkombination der Subtests ist die einfache Addition der ihnen zugeordneten Werte und es kann ein für beide Gruppen gemeinsamer Reliabilitätskoeffizient Ω geschätzt werden, dessen Wert nur von der mittleren Interkorrelation ($\bar{r} = 0{,}481$) und der Anzahl der Subtests ($m = 4$) abhängt:

$$\Omega = m * \bar{r} / [\, 1 + (m - 1) * \bar{r}\,] = 0{,}787$$

Eigentlich dient diese Formel nur der Berechnung von CRONBACHs α (sie findet sich bei NUNNALLY, 1967, S. 192 - 196). Bei Parallelität der Subtests jedoch entspricht Ω dem Koeffizienten α. Diese Tatsache haben wir genutzt, um auf einfache Weise Ω auszurechnen.

Normieren wir die Varianz der latenten Variablen ("Intellektuelle Leistungsfähigkeit") auf 1.0, ergibt sich die interne Validität λ_1 des CFT2 als: $\lambda_1 = \sqrt{0{,}787} = 0{,}887$. Die entsprechende Messfehler-Varianz $Var(\varepsilon_1)$ errechnet sich unter dieser Normierungsbedingung als: $Var(\varepsilon_1) = 1 - 0{,}787 = 0{,}213$.

3.1.2.2. Anstrengungsvermeidungstest (AVT)

Der AVT von ROLLETT & BARTRAM ist nach dem Messmodell von RASCH konstruiert. Indem die Autoren aber einen Koeffizienten der internen Konsistenz berechnen, schätzen sie die Güte ihres Tests in der klassischen, d.h. populationsabhängigen Weise. Diese Schätzung könnten wir zur Spezifikation unseres Messmodells übernehmen. Doch ist zu bedenken, dass sich das Messmodell von LISREL an den Prinzipien der klassischen Testtheorie bzw. common-factor-Theorie orientiert, denen auch die hier behandelten Homogenitätsdefinitionen folgen. Homogenität des Tests im Sinne von RASCH bedeutet nicht zwingend, dass die Homogenitätsvoraussetzungen gelten, die man der Berechnung des Konsistenzkoeffizienten zugrundelegen muss, und ein hoher Konsistenzkoeffizient bedeutet umgekehrt nicht notwendig Homogenität im Sinne der klassischen Testtheorie. So kann z.B. der Koeffizient α (bzw. Varianten oder Spezialfälle davon) auch dann einen ziemlich hohen Wert annehmen, wenn die Items nicht homogen (d.h. hier genauer: nicht mindestens essential-τ-äquivalent) sind.

Ferner kann daraus, dass der AVT über die Geschlechter hinweg RASCH-homogen ist, nicht gefolgert werden, dass die geschätzte Reliabilität in beiden Geschlechtergruppen gleich sein muss. Die Konsequenz aller dieser Ueberlegungen ist, dass wir prüfen müssen, ob in jeder Geschlechtergruppe die gleichen Parallelitäts-Bedingungen (d.h. Parallelitäts-Invarianz der Items) gelten. Erst bei Parallelitäts-Invarianz der Items kann ein gemeinsamer Koeffizient der internen Konsistenz berechnet werden.

Die Annahme der Parallelitäts-Invarianz wurde mit COFAMM überprüft. Analysiert wurden die Varianzen und Kovarianzen der 20 AVT-Items (sie wurden wie intervallskalierte Variable behandelt), die für Jungen und Mädchen getrennt in Tabelle 7 aufgeführt sind.

Tab. 7: Varianz-Kovarianz-Matrix der 20 Items des AVT, aufgegliedert nach Geschlecht

	Items	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Jungen	1	.253																			
	2	.055	.222																		
	3	.041	.029	.238																	
	4	.110	.029	.074	.238																
	5	.007	.054	.010	.010	.247															
	6	.069	.028	.043	.043	.017	.252														
	7	.000	.085	.001	.015	.040	.014	.222													
	8	.007	.043	.034	-.007	-.011	.030	.032	.251												
	9	-.007	.082	.024	.037	.110	.017	.068	.057	.247											
	10	.027	.028	.016	.043	-.010	.087	.083	.043	.017	.252										
	11	.034	.032	-.048	.020	.044	.057	.032	.004	.071	-.012	.251									
	12	.021	.020	.004	.032	.022	.047	.034	.058	.009	.047	.045	.217								
	13	.041	.034	-.006	.049	-.019	.030	.034	.029	-.019	.030	.056	.058	.200							
	14	.007	.017	.003	.030	.009	.006	-.010	.041	.051	.033	.027	-.008	-.004	.235						
	15	.007	.060	-.003	.065	.050	-.001	.073	-.019	.078	.081	.008	.029	.006	.052	.253					
	16	.041	.047	-.010	.018	.070	.020	.047	.010	.029	.033	.024	.046	.030	.012	.050	.156				
	17	.000	.058	.015	.042	-.001	.000	.085	-.010	.013	.096	.004	.020	.089	.058	.060	.020	.222			
	18	.027	.052	.026	.081	-.013	.013	.024	.064	.056	.068	-.018	.021	.034	.052	.064	-.006	.038	.222		
	19	-.014	.079	.009	-.032	.049	.030	.079	.033	.076	.030	.019	.023	.015	.027	.052	.037	.037	.004	.244	
	20	.034	.036	.000	.027	-.002	.004	.008	.015	-.016	.046	.001	.055	.016	.006	.037	-.007	.006	.024	.024	.242
Mädchen	1	.243																			
	2	.015	.233																		
	3	.046	.015	.243																	
	4	.064	.027	.067	.253																
	5	.015	.045	.058	.041	.233															
	6	.015	.048	.087	.064	.019	.248														
	7	.052	.033	-.009	.061	.048	.031	.179													
	8	.070	.047	.046	.058	-.026	.056	.065	.253												
	9	-.012	.089	.041	-.010	.046	.066	.043	.038	.219											
	10	.055	.061	.090	.116	.076	.070	.050	.064	.024	.253										
	11	.035	.038	-.020	.013	.009	.037	.049	.077	.042	.019	.246									
	12	.087	.007	.087	.073	.036	.058	.044	.065	-.015	.094	.036	.254								
	13	.064	.073	-.006	.003	.044	.015	.055	.035	.038	.049	.032	.015	.162							
	14	.067	.051	.049	.055	.094	.073	.075	.059	.064	.074	.048	.065	.084	.213						
	15	.049	.047	.038	.069	.061	.070	.010	.030	.015	.059	.045	.036	.032	.077	.201					
	16	.029	.038	.029	.039	.024	.052	.037	.057	.048	.028	.040	.007	.015	.051	.028	.171				
	17	.084	.008	.003	.018	-.021	-.025	-.002	.065	-.010	.022	.008	.015	.020	.023	.016	.017	.194			
	18	.038	.055	.064	.032	-.003	.046	.020	.014	.029	.028	-.002	.022	.026	.025	-.027	.030	-.016	.200		
	19	.046	.087	.012	.023	.000	.044	.020	.075	.070	.032	.096	.029	.038	.035	.052	.073	.003	.035	.243	
	20	.044	.070	.029	.019	.041	.050	.021	.008	.054	.052	.025	.000	.087	.101	.081	.010	.027	.006	.015	.207

Der statistische Vergleich der in Tab. 7 aufgeführten Varianz-Kovarianz-Matrizen unter der Parallelitäts-Bedingung ergab einen Wert von $\chi^2 = 530{,}366$, der bei 417 Freiheitsgraden signifikant ist ($P < 0{,}001$). Die Hypothese der Parallelitäts-Invarianz der AVT-Items ist also zu verwerfen und für jede Geschlechtergruppe ist die interne Reliabilität separat zu schätzen.

Innerhalb jeder Geschlechtergruppe kann die Parallelitäts-Bedingung aber immer noch gelten. Wir haben deshalb nicht sofort die nächstschwächere Homogenitätsbedingung (τ-Aequivalenz) vergleichend beurteilt, sondern erst noch die Parallelitäts-Bedingung in jeder Gruppe separat überprüft. Die Ergebnisse dieser separaten Tests stehen in den Tabellen 8 (für die Jungen) und 9 (für die Mädchen). Es wurde hier mit dem Programm LISREL gerechnet (mit dem Programm COFAMM geht das selbstverständlich auch).

Ueberprüfung der Parallelitäts-Bedingung bei den Jungen:

Tab. 8: Analyse der Varianz-Kovarianz-Matrix von Tab. 7 (n = 74 Jungen) unter zwei verschiedenen Homogenitätsbedingungen

Hypothesen	df	χ^2	P
parallel	208	269,677	0,0026
kongenerisch	170	240,569	0,0003

Die Einzelanalyse der Daten unter der Hypothese kongenerischer Items (hier auch genannt: Ein-Faktor-Hypothese) brachte einen signifikanten χ^2-Wert ($P < 0{,}001$). Zunächst ist man geneigt, die Ein-Faktor-Hypothese als nicht haltbar im signifikanzstatistischen Sinne abzulehnen bzw. mehr als einen Faktor zur Erklärung der Daten heranzuziehen. Da aber

die restriktivere Parallelitäts-Hypothese statistisch nicht verworfen werden kann ($P > 0{,}001$), genügt es hier sicher nicht zu fragen, wieviele Faktoren im statistischen Sinne signifikant sind, sondern wieviele Faktoren man im vorliegenden Einzelfall als angemessen ansehen bzw. als sinnvoll beibehalten soll.

Um einen weiteren Anhaltspunkt für die Anzahl der Faktoren zu gewinnen, wurde die Analyse unter der Ein-Faktor-Hypothese mit der kanonischen Faktorenanalyse von RAO wiederholt und nach einem Knick im Eigenwertdiagramm gesucht. LISREL liefert Maximum-Likelihood-Lösungen und das Verfahren von RAO kann man als Spezialfall der Maximum-Likelihood-Methode ansehen.

Nach statistischen Gesichtspunkten brachte diese Wiederholung das (nahezu) gleiche Ergebnis: einen Wert von $\chi^2 = 240{,}616$ bei 170 Freiheitsgraden (vgl. in Tab. 8). Die ergänzende Betrachtung des Diagramms der 20 Eigenwerte gab einen klareren Anhaltspunkt für die Anzahl der Faktoren: Der Eigenwertverlauf ist so, dass man praktisch nur einen einzigen Faktor als relevant ansehen kann.

Nun konnte - wir kehren jetzt zur LISREL-Analyse (Tab. 8) zurück - die Parallelitäts-Hypothese gegen die Ein-Faktor-Hypothese getestet werden. Der χ^2-Test einer Anpassungsverschlechterung durch das Parallelitätsmodell ergab ein $\chi^2 = 29{,}108$ (= 269,677 - 240,569), das bei 38 (= 208 - 170) Freiheitsgraden einem $P > 0{,}001$ (genauer: $P \simeq 0{,}80$) entspricht. Die Differenz der χ^2-Werte ist demnach statistisch nicht signifikant: Die Parallelitäts-Hypothese stellt also eine gegenüber der allgemeineren Ein-Faktor-Hypothese annehmbare einfachere Messmodell-Alternative dar.

Wir gehen deshalb davon aus, dass die Itemwerte nicht gewichtet zu werden brauchen, um maximal reliable AVT-Werte zu erzielen. Da Item-Varianzen und -Kovarianzen statistisch

gesehen nicht signifikant voneinander abweichen, kann die Reliabilität nach der KUDER-RICHARDSON-Formel 21 (s. LIENERT, 1961, S. 222) geschätzt werden. Diese Schätzung ergab ein $r_{tt} = 0,747$.

Bei Standardisierung der latenten Variablen im Strukturmodell (Varianz = 1,0) berechnet sich die interne Validität λ_8 des AVT als: $\lambda_8 = \sqrt{0,747} = 0,864$ und die entsprechende Messfehler-Varianz $Var(\varepsilon_8)$ des AVT als: $Var(\varepsilon_8) = 1 - 0,747 = 0,253$.

Ueberprüfung der Parallelitäts-Bedingung bei den Mädchen:

Tab. 9: Analyse der Varianz-Kovarianz-Matrix von Tab. 7 (n = 70 Mädchen) unter zwei verschiedenen Homogenitätsbedingungen

Hypothesen	df	χ^2	P
parallel	208	257,008	0,0116
kongenerisch	170	222,376	0,0043

Die statistische Analyse der Daten unter der Hypothese kongenerischer Items brachte keinen signifikanten χ^2-Wert ($P > 0,001$). Das Ein-Faktor-Modell reproduziert demnach die Daten hinreichend gut.

Der χ^2-Test einer Anpassungsverschlechterung durch das Parallelitätsmodell ergab ein $\chi^2 = 34,632$ (= 257,008 - 222,376), das bei 38 (= 208 - 170) Freiheitsgraden einem $P > 0,001$ (genauer: $0,50 < P < 0,60$) entspricht. Die χ^2-Werte der beiden auf die Daten passenden Homogenitätsmodelle unterscheiden sich also nicht signifikant voneinander. Daraus lässt sich ableiten, dass der AVT auch bei den Mädchen aus

lauter parallelen Items besteht.

Dementsprechend ist auch hier die einfache Summe der Itemwerte gleichzeitig die maximal reliable Linearkombination. Die Schätzung der internen Reliabilität nach der KUDER-RICHARDSON-Formel 21 ergab ein $r_{tt} = 0,812$.

Bei Standardisierung der latenten Variablen im Strukturmodell (Varianz = 1,0) berechnet sich bei den Mädchen die interne Validität λ_8 des AVT als: $\lambda_8 = \sqrt{0,812} = 0,901$ und die entsprechende Messfehler-Varianz $Var(\varepsilon_8)$ des AVT als: $Var(\varepsilon_8) = 1 - 0,812 = 0,188$.

Folgerungen

Eingangs leiteten wir aus dem statistischen Ergebnis des Tests auf Parallelitäts-Invarianz die Entscheidung ab, keinen für beide Geschlechter gemeinsamen Koeffizienten der internen Reliabilität bzw. Validität zu schätzen. Doch würden wir den von den Testautoren geschätzten Reliabilitätswert übernehmen, dann läge die Schätzung der internen Validität des AVT mit $\sqrt{0,80} = 0,894$ zwischen den von uns errechneten Werten (0,864 für die Jungen; 0,901 für die Mädchen). Man kann sich fragen, ob man die numerischen Abstände zum mittleren Wert auch praktisch für bedeutsam halten soll oder nicht.

Wir halten diese Unterschiede für vernachlässigbar. Da die Parallelitäts-Bedingung in beiden Geschlechtergruppen ohnehin erfüllt ist, macht man sicher keinen die Schätzung der Strukturparameter grob verzerrenden Fehler, wenn wir die interne Validität in beiden Geschlechtergruppen auf den mittleren Wert $\lambda_8 = 0,894$ festlegen und die Messfehler-Varianz dementsprechend auf den Wert $Var(\varepsilon_8) = 1 - 0,8 = 0,2$.

3.1.2.3. Index der sozialen Herkunft (BRSS)

BAUER macht zur Reliabilität seines Indexes keine numerischen Angaben. Da uns ebenfalls Informationen zu den Komponenten des BRSS (Schulbildung von Vater und Mutter, Berufsposition des Vaters) fehlten, konnte auch anhand unserer eigenen Daten keine direkte Schätzung der internen Konsistenz, wie beim CFT2 und AVT geschehen, vorgenommen werden. Daher versuchten wir auf indirektem Wege zu Schätzungen der internen Reliabilität und Validität zu gelangen: über eine Faktorenanalyse der von BAUER mitgeteilten Korrelationen des BRSS mit vergleichbaren Indizes der sozialen Herkunft (um im folgenden Begriffsverwirrungen zu vermeiden: Im faktorenanalytischen Zusammenhang wollen wir "Indizes", die dieselbe latente Variable erfassen sollen, als "Indikatoren" bezeichnen).

Normalerweise liefert die Faktorenanalyse eine untere Grenze für die Reliabilität eines jeden Indikators. Wenn jedoch die Annahmen gelten, dass die betrachteten Indikatoren

1. nur einen einzigen gemeinsamen Faktor,
2. gleiche Faktorenladungen (bzw. gleiche Interkorrelationen),
3. nicht-korrelierende Messfehler

haben, dann schätzt die Faktorenanalyse die Reliabilität dieser Indikatoren genau und deren (standardisierte) Faktorenladungen kann man als interne Validitäten interpretieren.

BAUER führt Interkorrelationen des BRSS mit 6 weiteren Indikatoren der sozialen Herkunft (BAUER: "Schichtmasse", 1972, S. 7) auf. Für unsere Faktorenanalyse liessen sich jedoch nur zwei weitere Indikatoren heranziehen: die SSE-Skala von KLEINING & MOORE (1968) und ein in Anlehnung an PEISERT & DAHRENDORF (1967) entwickelter Indikator, den wir hier abgekürzt "PDS" nennen wollen.

Von BAUER übernehmen wir folgende Interkorrelationen:

	BRSS	PDS
SSE	0,81	0,83
BRSS		0,82

Die übrigen von BAUER verwendeten Indikatoren waren entweder mit dem BRSS inhaltlich nicht vergleichbar (z.B. eine reine Prestige-Skala, obschon sie mit dem BRSS hoch zu 0,85 korrelierte) oder aber sie erschienen, was BAUER selber auch betont, (experimentell) nicht unabhängig vom BRSS gewonnen. Die hier herangezogenen drei Indikatoren dagegen können wir als unabhängige Mehrfachmessungen der sozialen Herkunft auffassen.

Weiterhin nehmen wir an, dass die Messfehler-Varianz des BRSS unabhängig davon ist, ob die in ihm verarbeiteten Angaben zur Schulbildung, usw. von den Eltern der Mädchen oder von den Eltern der Jungen gemacht wurden. Von daher erscheint es nicht notwendig, interne Reliabilität wie Validität für jede Geschlechtergruppe separat zu schätzen.

Angesichts der oben vorliegenden Korrelationsstruktur unabhängig messender Indikatoren können wir davon ausgehen, dass die Annahmen eines einzigen gemeinsamen Faktors, gleicher Faktorenladungen und nicht-korrelierender Messfehler annähernd richtig sind. Unter diesen Bedingungen lassen sich die internen Validitäten β_i direkt aus den beobachteten Korrelationen errechnen. Das in Abbildung 10 in Form eines Pfaddiagramms dargestellte Modell ist genau identifizierbar, da es drei Gleichungen für drei freie, unbekannte Parameter impliziert.

In diesem Untersuchungszusammenhang erscheint die Betrachtung von BRSS als "Wirkungsindikator" angemessen (vgl. dazu unsere Ueberlegungen auf S. 56); denn Analyseeinheiten sind hier

nicht die einzelnen Komponenten "Schulbildung", "Berufsposition", usw., sondern bereits vollzogene Zusammenfassungen dieser Komponenten: Indizes.

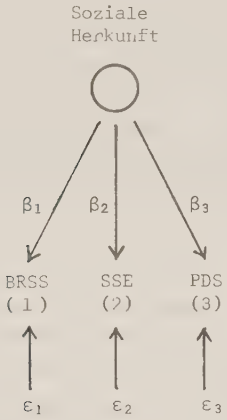

Abb. 10: Ein-Faktor-Modell zur Erklärung der Korrelationen zwischen drei Indikatoren der sozialen Herkunft (Erläuterung zu den Indikatoren s. Text).
Wobei: β = standardisierte Faktorenladungen
ε = Messfehler-Variable

Identifizierende Gleichungen (Varianz des Faktors "Soziale Herkunft" = 1):

(1) $r_{12} = 0,81 = \beta_1 * \beta_2$

(2) $r_{13} = 0,82 = \beta_1 * \beta_3$

(3) $r_{23} = 0,83 = \beta_2 * \beta_3$

Aus den Gleichungen (2) und (3) folgt:

$$\beta_2 = \beta_1 * \frac{r_{23}}{r_{13}}$$

Dieses Ergebnis in (1) eingesetzt, ergibt:

$$r_{12} = \beta_1^2 * \frac{r_{23}}{r_{13}}$$

$$\beta_1^2 = \frac{r_{13} * r_{12}}{r_{23}}$$

$$\beta_1 = \sqrt{\frac{0,82 * 0,81}{0,83}} = 0,895$$

In Ermangelung anderer Daten interpretieren wir β_1 als internen Validitätskoeffizienten des BRSS, d.h. wir fixieren λ_9 im LISREL-Modell auf 0,895. Die entsprechende Messfehler-Varianz des BRSS beträgt: $Var(\delta) = 1 - (0,895)^2 = 0,200$.

3.1.3. Linearität und Additivität der Indikatoren-Beziehungen

Linearität

Für Indikatoren verschiedener latenter Variablen wurde mit der Prozedur BREAKDOWN aus dem Programmsystem SPSS paarweise überprüft, ob eine nicht-lineare Regression statistisch signifikant besser als eine lineare ist. Je Geschlecht wurden 33 Paarbeziehungen von unabhängigen und abhängigen Indikatoren überprüft. Die Unterscheidung in "unabhängig" und "abhängig" geschah entsprechend der Kausalanordnung der zugehörigen latenten Variablen.

Ergebnis: Bis auf eine Ausnahme konnten sämtliche Linearitätsannahmen nicht widerlegt werden. Die Ausnahme war die Regression des Lehrer-Sympathie-Index (LESY) auf den Index der sozialen Herkunft (BRSS) bei den Mädchen.

Additivität

Für Indikatoren verschiedener latenter Variablen wurde mit der Prozedur REGRESSION aus dem Programmsystem SPSS überprüft, ob in einer Regressionsanalyse zusätzlich zu den additiven Effekten der unabhängigen Indikatoren auch nicht-additive Effekte vorhanden waren. Dabei liessen wir nur multiplikative Interaktionen zu. Je nach Anzahl der unabhängigen Indikatoren in der Regressionsgleichung wurden multiplikative Beziehungen erster, zweiter, dritter und vierter Ordnung verwendet, um alle möglichen Interaktionseffekte darzustellen.

Ergebnis: Die Ueberprüfung von insgesamt 142 Interaktionsvariablen erster Ordnung, von 116 Interaktionsvariablen zweiter Ordnung, 48 Interaktionsvariablen dritter Ordnung und 8 Interaktionsvariablen vierter Ordnung ergab in nur zwei Fällen bei den Mädchen und in nur einem Fall bei den Jungen eine Widerlegung der Additivitätsannahmen. In der Vielzahl der Ueberprüfungen erschienen uns diese wenigen (unsystematischen) Widerlegungen eher per Zufall signifikant.

Folgerungen

Wenn wir eine möglichst einfache Art der Beschreibung der Beziehungen unter den Indikatoren verschiedener latenter Variablen anstreben, eine Beschreibung, die zugleich vollständig und hinreichend ist, dann sollten die hier aufgedeckten nicht-linearen und nicht-additiven Beziehungen nur berücksichtigt werden, wenn das einfachere Gesamtmodell mit den durchgehend linear-additiven Indikatoren-Beziehungen die Daten nicht genau genug reproduziert. In diesem Lichte erscheinen die wenigen Abweichungen von der Annahme der Linearität und Additivität minimal und vernachlässigbar: Die wenigen aufgetretenen "Störungen" lassen nicht erwarten, dass ein vollständig rekursiv spezifiziertes Strukturgleichungssystem an die Daten unangepasst wäre.

3.1.4. Test der Messfehler-Hypothesen

Wir analysierten die Indikator-Korrelationen beider Geschlechtergruppen (s. Tab. 10) unter folgenden Messfehler-Hypothesen H_1 - H_3:

H_1 : λ_s als freier Parameter

H_2 : $\lambda_s = 0$, d.h. kein systematischer Messfehler

H_3 : $\lambda_s = 0$, $\lambda_2 = \lambda_3$, $\lambda_4 = \lambda_5$, $\lambda_6 = \lambda_7$,
$Var(\varepsilon_2) = Var(\varepsilon_3)$, $Var(\varepsilon_4) = Var(\varepsilon_5)$, $Var(\varepsilon_6) = Var(\varepsilon_7)$,
d.h. Parallelität der multiplen Indikatoren

Tab. 10: Korrelationen der Indikatoren, aufgegliedert nach Geschlecht: Jungen unterhalb, Mädchen oberhalb der Hauptdiagonalen.

	BRSS	CFT2	MATH	SPRA	MUNT	VUNT	LUNT	LESY	AVT
BRSS		.531	.601	.536	.223	.214	.306	.353	.512
CFT2	.258		.698	.594	.120	.063	.231	.325	.418
MATH	.456	.514		.821	.203	.205	.269	.422	.559
SPRA	.555	.564	.731		.284	.267	.439	.472	.524
MUNT	.172	.030	.034	.162		.592	.634	.529	.379
VUNT	.167	.130	.296	.191	.490		.616	.375	.282
LUNT	.094	.015	.089	.059	.106	.070		.668	.479
LESY	.048	.070	.247	.171	.093	.143	.500		.638
AVT	.339	.110	.457	.474	.140	.207	.342	.346	

Wobei: BRSS = Index der sozialen Herkunft von BAUER, CFT2 = Grundintelligenztest-Skala 2 von WEISS, MATH und SPRA = Kurszugehörigkeiten in Mathematik und Sprachen, MUNT und VUNT = Marburger Unterstützungsskalen: Mutter- und Vaterversion, LUNT = Lehrer-Unterstützungsskala von HEINRICH, LESY = Lehrer-Sympathie-Index von BRUSTEN und HURRELMANN, AVT = Anstrengungsvermeidungstest von ROLLETT und BARTRAM

An dieser Stelle erscheint es angebracht, auch gleich die arithmetischen Mittel und Standardabweichungen der Indikatoren anzuführen (s. Tab. 11). Von Nutzen sind diese Informationen beispielsweise, wenn man aus der vorstehenden Korrelationsmatrix eine Varianz-Kovarianz-Matrix ableiten will.

Tab. 11: Arithmetische Mittel ($\bar{x}$) und Standardabweichungen (s) der Indikatoren, aufgegliedert nach Geschlecht.

		BRSS	CFT2	MATH	SPRA	MUNT	VUNT	LUNT	LESY	AVT
Jungen	$\bar{x}$	2,70	30,42	2,00	1,99	52,90	52,57	41,68	-0,16	10,77
	s	1,36	5,77	0,79	0,80	8,68	10,88	8,75	0,84	4,01
Mädchen	$\bar{x}$	3,36	32,11	1,76	1,67	51,54	49,29	44,11	0,17	12,11
	s	1,75	6,57	0,82	0,74	9,40	13,58	12,04	0,84	4,41

Bei stets uneingeschränktem, d.h. vollständig rekursivem Strukturmodell wurden die Hypothesen H_1 - H_3 einzeln gegen die Alternativhypothese getestet, dass Σ nicht restringiert ist. Die Ergebnisse für die Jungen finden sich in Tab. 12 die für die Mädchen in Tab. 14.

Zum Test der Messfehler-Hypothesen wurden dann die χ^2-Werte aus den Einzelanalysen miteinander verglichen. Die Resultate dieses Vergleichs stehen in den Tabellen 13 (Jungen) und 15 (Mädchen). Der Zufallsfehler-Test entspricht dem Test von H_2 (= Nullhypothese) gegen H_1 (= Alternativhypothese), der Parallelitäts-Test dem Test von H_3 (= Nullhypothese) gegen H_2 (= Alternativhypothese).

Ergebnisse: Jungen

Tab. 12: Analyse der Korrelationen von Tab. 10 (Gruppe: Jungen) unter der Hypothese eines systematischen Messfehlers (H_1), zufälliger Messfehler (H_2) und paralleler Indikatoren (H_3).

Hypothesen	df	χ^2	P
H_1	12	14,9617	0,2435
H_2	13	15,3160	0,2880
H_3	19	17,8641	0,5315

Tab. 13: Test der Zufallsfehler-Hypothese (H_2 gegen H_1) und der Parallelitäts-Hypothese (H_3 gegen H_2): Vergleich der χ^2-Werte von Tab. 12

Tests	df	χ^2	P
H_2 gegen H_1	1	0,3543	> 0,001
H_3 gegen H_2	6	2,5481	> 0,001

Ergebnisse: Mädchen

Tab. 14: Analyse der Korrelationen von Tab. 10 (Gruppe: Mädchen) unter der Hypothese eines systematischen Messfehlers (H_1), zufälliger Messfehler (H_2) und paralleler Indikatoren (H_3).

Hypothesen	df	χ^2	P
H_1	12	17,2646	0,1399
H_2	13	27,4315	0,0109
H_3	19	31,3465	0,0370

Tab. 15: Test der Zufallsfehler-Hypothese (H_2 gegen H_1) und der Parallelitäts-Hypothese (H_3 gegen H_2): Vergleich der χ^2-Werte von Tab. 14

Tests	df	χ^2	P
H_2 gegen H_1	1	7,1669	> 0,001
H_3 gegen H_2	6	3,9150	> 0,001

Folgerungen

Bei allen Tests erhielten wir nicht-signifikante χ^2-Werte ($P > 0,001$). Die Hypothese zufälliger Messfehler und paralleler Indikatoren liess sich also in beiden Geschlechtergruppen statistisch nicht widerlegen.

In beiden Geschlechtergruppen können wir also Parallelität der multiplen Indikatoren unterstellen. Zur Prüfung der eigentlichen Theorie lässt sich dort, wo wir es wollen, mehr als ein Indikator je latenter Variable verwenden.

Weiterhin besagt dieses Ergebnis, dass die Schätzung der Faktorenladungen nicht wesentlich durch konzeptfremde Indikatoren beeinflusst worden ist: Die latenten Variablen sind eindeutig von den ihnen zugeordneten Indikatoren her zu interpretieren. Dieses Ergebnis ist wichtig; denn starke strukturelle Abhängigkeiten der Faktorenladungen erschweren nicht nur die beabsichtigte eindeutige Interpretation der latenten Variablen innerhalb des Modells, sondern machen möglicherweise auch die zum Gruppenvergleich (Jungen versus Mädchen) geforderte Interpretationsinvarianz der latenten Variablen zunichte.

Berücksichtigt man auch noch den Sachverhalt, dass zwischen den Messmodell-Parametern der Einzelindikatoren keine Geschlechtsunterschiede bestehen, dann kann man aus den hier in beiden Geschlechtergruppen nicht-signifikanten Testergebnissen zugleich auch ableiten, dass die (separat betrachteten) Messmodelle populationsunabhängig, hier genauer: geschlechtsunabhängig definiert sind. Wir können davon ausgehen, dass die Messmodelle strukturell homogen sind und damit die Voraussetzung zum Gruppenvergleich der Strukturparameter gegeben ist.

3.2. Test der Alternativen zur Vermittlungshypothese und Test der Zielsetzung der Förderstufe

Zunächst analysierten wir die Korrelationen der Indikatoren in jeder Geschlechtergruppe unter der Bedingung H_4, d.h. unter Berücksichtigung der Parallelitäts-Restriktionen wurde die Hypothese: $\gamma_3 = \gamma_4 = \gamma_5 = \beta_9 = 0$ gegen die Alternativhypothese getestet, dass Σ nicht restringiert ist. Die Ergebnisse dieser Analysen finden sich in Tabelle 16.

Zum statistischen Test der ersten Alternative zur Vermittlungshypothese wurden dann die χ^2-Werte der Analysen unter H_4 und H_3 (= Vermittlungshypothese) miteinander verglichen. Das Resultat dieses Vergleichs steht in Tabelle 17.

Tab. 16: Analyse der Korrelationen von Tab. 10 unter der Nullhypothese H_4: $\gamma_3 = \gamma_4 = \gamma_5 = \beta_9 = 0$

Vpn	df	χ^2	P
Jungen	23	18,5862	0,7250
Mädchen	23	37,0155	0,0324

Tab. 17: Test der Nullhypothese H_4: $\gamma_3 = \gamma_4 = \gamma_5 = \beta_9 = 0$ gegen die Alternativhypothese H_3: freie Strukturparameter

Vpn	df	χ^2	P
Jungen	4	0,7221	> 0,001
Mädchen	4	5,6690	> 0,001

Der Test der Nullhypothese H_4 führt in beiden Geschlechtergruppen zu einem nicht-signifikanten χ^2-Wert ($P > 0{,}001$). Wir akzeptieren dementsprechend die Nullhypothese und verwerfen die Alternativhypothese.

Dieses Testergebnis ist gleichbedeutend mit der Widerlegung der Vermittlungshypothese BORCHERT & MASENDORFs, die hier der Alternativhypothese entspricht.

Unter diesen Umständen ist es sinnvoll, auch den zweiten, schärferen Kontrast zur Vermittlungshypothese zu testen, also auch die zusätzlich in den Strukturparametern β_2, β_3 und β_4 formulierte Nullhypothese. Bei gleichbleibender Darstellung der Alternativhypothese durch das uneingeschränkte Strukturmodell können wir auch hier wieder die χ^2-Werte der Einzelanalysen miteinander vergleichen. Zuvor aber muss die neue Nullhypothese (= H_5), wiederum unter Berücksichtigung der Parallelitäts-Bedingungen multipler Indikatoren, gegen die allgemeinste Alternativhypothese getestet werden, dass Σ uneingeschränkt ist. Die Ergebnisse dieser Analyse zeigt Tabelle 18, die Ergebnisse der χ^2-Wert-Vergleiche Tabelle 19.

Tab. 18: Analyse der Korrelationen von Tab. 10 unter der Nullhypothese H_5: $\gamma_3 = \gamma_4 = \gamma_5 = \beta_2 = \beta_3 = \beta_4 = \beta_9 = 0$

Vpn	df	χ^2	P
Jungen	26	26,4262	0,4399
Mädchen	26	38,7278	0,0518

Tab. 19: Test der Nullhypothese H_5: $\gamma_3 = \gamma_4 = \gamma_5 = \beta_2 = \beta_3 = \beta_4 = \beta_9 = 0$ gegen die Alternativhypothese H_3: freie Strukturparameter

Vpn	df	χ^2	P
Jungen	7	8,5621	> 0,001
Mädchen	7	7,3813	> 0,001

Auch der Test von Nullhypothese H_5, eines Spezialfalls der Nullhypothese H_4 ergibt in beiden Geschlechtergruppen einen nicht-signifikanten χ^2-Wert ($P > 0,001$), d.h. auch hier ist die Alternativhypothese generell zugunsten der Nullhypothese zu verwerfen.

Dieses Testergebnis ist ebenfalls gleichbedeutend mit der Widerlegung der Vermittlungshypothese von BORCHERT & MASENDORF. Unter der Voraussetzung, dass Lehrerurteil (z.B. Kausalattribuierungen) und Lehrerverhalten (z.B. Unterstützung) übereinstimmen, widerspricht dieses Ergebnis auch der Hypothese SCHWARZERs, das Lehrer(urteils-)verhalten werde direkt von der sozialen Herkunft des Schülers und seiner Testintelligenz beeinflusst.

In der Hypothese von der entscheidenden vermittelnden Variablen "Kurszugehörigkeit" ist impliziert, dass die Kurszugehörigkeit nicht nur auf Eltern- und Lehrer-Unterstützung, sondern auch auf die Anstrengungsbereitschaft direkt wirkt, d.h. es muss $\beta_7 \neq 0$ sein. Die Annahme der Nullhypothese H_5 selber gibt jedoch noch keine Auskunft darüber, ob das der Fall ist. Zu fragen ist daher, ob über die unter H_5 vorgegebenen Nullrestriktionen hinaus die Zusatzrestriktion $\beta_7 = 0$ die Modellanpassung wesentlich verschlechtert oder nicht. Zugleich entspricht der Test der Zusatzrestriktion $\beta_7 = 0$ gegen H_5 der

Ueberprüfung, ob die Förderstufe ihr Ziel "Fördern statt Auslesen", bezogen auf die Anstrengungsbereitschaft, erreicht oder nicht. Die Ergebnisse der Analyse der Korrelationen unter der Zusatzrestriktion zeigt Tabelle 20, den Test der Zusatzrestriktion Tabelle 21.

Tab. 20: Analyse der Korrelationen von Tab. 10 unter der Zusatzrestriktion $\beta_7 = 0$, d.h. unter der Nullhypothese: $\gamma_3 = \gamma_4 = \gamma_5 = \beta_2 = \beta_3 = \beta_4 = \beta_7 = \beta_9 = 0$

Vpn	df	χ^2	P
Jungen	27	39,4156	0,0581
Mädchen	27	47,0833	0,0097

Tab. 21: Test der Zusatzrestriktion $\beta_7 = 0$ gegen die Alternativhypothese H_5

Vpn	df	χ^2	P
Jungen	1	12,9894	$P < 0,001$
Mädchen	1	8,3555	$0,001 < P < 0,010$

Der Test der Zusatzrestriktion $\beta_7 = 0$ ergibt bei den Jungen einen signifikanten, bei den Mädchen dagegen einen knapp nichtsignifikanten χ^2-Wert. Demnach wäre die Alternativhypothese H_5 bei den Jungen beizubehalten, bei den Mädchen aber abzulehnen. Betrachtet man aber die Restkorrelations-Matrix bei den Mädchen genauer, dann zeigt sich, dass mehrere Restkorrelationen gerade wegen der Zusatzrestriktion ziemlich gross geworden sind (z.B. von 0,006 unter H_5 auf jetzt 0,183 oder von 0,120

unter H_5 auf jetzt 0,249). Unter diesen Umständen erscheint uns die Beibehaltung von $\beta_7 = 0$ auch bei den Mädchen nicht vertretbar.

Damit bestätigt dieser Test in beiden Geschlechtergruppen über die allgemeine Vermutung hinaus auch noch die speziellere Erwartung, dass der Einfluss der sozialen Herkunft auf die schulische Anstrengungsbereitschaft entscheidend durch die Kurszugehörigkeit vermittelt wird.

Im Zusammenhang des Testens von Hypothesen ist es eigentlich nicht notwendig, Genaueres über die Schätzwerte der (frei gebliebenen) Parameter zu wissen. Dennoch seien diese - der Vollständigkeit halber und für nachfolgende Explorationen - mitgeteilt (s. Tab. 22):

b. 22: Parameterschätzungen unter den Hypothesen H_1 - H_5, standardisierte Lösungen.
* = während der Schätzung auf "bekannte" Werte fixierte Parameter; a - f = während der Schätzung auf gleiche, aber unbekannte Werte fixierte Parameter (Erläuterungen zu den Hypothesen H_1 - H_5 siehe S.7c)

	unter H_1		unter H_2		unter H_3		unter H_4		unter H_5	
arameter	Jungen	Mädchen	Jungen	Mädchen	Jungen	Mädchen	Jungen	Mädchen	Jungen	Mädchen
1	.887*	.887*	.887*	.887*	.887*	.887*	.887*	.887*	.887*	.887*
2	.813	.958	.813	.960	.855a	.907a	.857a	.904a	.861a	.904a
3	.899	.858	.900	.856	.855a	.907a	.857a	.904a	.861a	.904a
4	.560	.792	.582	.814	.700b	.770b	.700b	.770b	.700b	.770b
5	.875	.753	.843*	.727*	.700b	.770b	.700b	.770b	.700b	.770b
5	-.092	.599	.000	.000	.000*	.000*	.000*	.000*	.000*	.000*
6	.751	.330	.701	.838	.707c	.817c	.707c	.826c	.706c	.826c
7	.687*	.946*	.714*	.797*	.707c	.817c	.707c	.826c	.706c	.826c
8	.895*	.895*	.895*	.895*	.895*	.895*	.895*	.895*	.895*	.895*
9	.895	.895	.895	.895	.895	.895	.895	.895	.895	.895
1	.549	.624	.549	.624	.555	.606	.551	.585	.516*	.586*
2	-.098	-.320	-.097	-.385	-.116	-.476	-.172	-.394	.000*	.000*
3	-.177	.111	-.197	.225	-.222	.178	-.146	.150	.000*	.000*
4	-.542	-.112	-.536	-.279	-.541	-.214	-.496	.083	.000	.000
5	.344	.353	.331	.416	.291	.552	.417	.665	.289	.329
6	.352	.234	.405	.027	.451	136	.284	.160	.164	.295
7	.967	.280	.949	.353	.953	.228	.864	.333	.475	.409
8	.188	.539	.132	.827	.150	.790	.143*	.755*	.173*	.734*
9	-.006	-.025	.024	-.648	.020	-.623	.000	.000	.000	.000
10	.391	.553	.378	1.125	.362	1.111	.388	.516	.446	.509
1	.326	.669	.326	.669	.326	.669	.332	.665	.326	.669
2	.492	.283	.492	.282	.481	.296	.470*	.330*	.496*	.325*
3	.029	.290	.050	.281	.116	.249	.000*	.000*	.000*	.000*
4	-.091	.006	-.127	.015	-.154	-.016	.000*	.000*	.000*	.000*
5	-.098	.295	-.086	.283	-.077	.321	.000	.000	.000	.000
ar(ζ_1)	.894	.552	.894	.552	.894	.552	.890	.558	.894	.552
ar(ζ_2)	.281	.293	.281	.295	.288	.305	.303	.293	.321	.296
ar(ζ_3)	.907	.854	.905	.841	.901	.795	.898	.824	.916	.892
ar(ζ_4)	.901	.501	.914	.176	.896	.188	.918	.228	.927	.231
ar(ζ_5)	.321	.281	.328	.209	.325	.206	.342	.351	.485	.349
ar(ε_1)	.213*	.213*	.213*	.213*	.213*	.213*	.213*	.213*	.213*	.213*
ar(ε_2)	.339	.083	.339	.079	.268d	.178d	.265d	.182d	.258d	.182d
ar(ε_3)	.192	.264	.191	.267	.268d	.178d	.265d	.182d	.258d	.182d
ar(ε_4)	.687	.373	.661	.338	.510e	.408e	.510e	.408e	.510e	.408e
ar(ε_5)	.235	.433	.289	.471	.510e	.408e	.510e	.408e	.510e	.408e
ar(ε_6)	.461	.284	.508	.298	.500f	.332f	.500f	.318f	.501f	.318f
ar(ε_7)	.529*	.106*	.490*	.365*	.500*	.332*	.500*	.318*	.501*	.318*
ar(ε_8)	.200*	.200*	.200*	.200*	.200*	.200*	.200*	.200*	.200*	.200*
ar(δ)	.200	.200	.200	.200	.200	.200	.200	.200	.200	.200

Zusammenfassend: Die Verwerfung von $\beta_7 = 0$, nachdem H_5 als Nullhypothese nicht verworfen werden konnte, führt zur Feststellung, dass das von der Förderstufe selbstgesteckte Ziel "Fördern statt Auslesen", bezogen auf die Anstrengungsbereitschaft, nicht erreicht wird. Die Verwerfung von $\beta_7 = 0$ ist hier äquivalent dem Nicht-Verwerfen-Können von H_5, und das Nicht-Verwerfen von H_5 gegenüber H_3, dem vollständig rekursiven Strukturmodell, entspricht der Widerlegung der Vermittlungshypothese von BORCHERT & MASENDORF.

Die hypothesentestende Analyse ist hiermit beendet. Das Testergebnis ist reproduzierbar und diese Kreuzvalidierung erhöht das Vertrauen in die Stabilität der postulierten Kausalbeziehungen gegenüber Stichprobenfehlern.

3.3. Ergänzende Bemerkungen

Ergänzend zur hypothesentestenden Analyse wurde noch explorativ untersucht (hier nicht weiter dargestellt), ob zusammen mit den bereits auf die Daten passenden Nullrestriktionen ("passend" von den Hypothesentests her), noch weitere Nullrestriktionen vereinbar waren.

Das Ergebnis in Kürze: Bei den Mädchen liess sich die unter H_5 vorgegebene Kausalstruktur nicht weiter vereinfachen, d.h. jede zusätzliche Nullrestriktion führte zu einer signifikanten Unverträglichkeit der empirischen mit den reproduzierten Korrelationen. Bei den Jungen dagegen lag eine Vereinfachungsmöglichkeit vor: Zu den unter H_5 schon vorgegebenen Nullrestriktionen liess sich ohne bedeutsame Anpassungsverschlechterung zusätzlich noch β_8 auf Null festlegen.

Diese Ergebnis-Feststellung impliziert, dass auch die Kausalbeziehungen von der Eltern- zur Lehrer-Unterstützung und von

der Kurszugehörigkeit zur Lehrer-Unterstützung in beiden Geschlechtergruppen "nicht verschwanden", d.h. auch die Zusatzrestriktion $\beta_6 = \beta_8 = 0$ war mit den Daten nicht vereinbar. Bezogen auf den Gesamtzusammenhang ergibt sich so ein kausalstruktureller Sachverhalt, den man im Kontext der vorangegangenen hypothesentestenden Verwerfung von $\beta_7 = 0$ gegenüber H_5, den Ausdruck "Fördern statt Auslesen" gegenzeichnend, als "Unverträglichkeit von Fördern und Auslesen" charakterisieren könnte, d.h. bei gegebener Auslese kann die Förderung (durch den Lehrer) nicht unabhängig von der Kurszugehörigkeit des Schülers erfolgen.

Das explorative Ergebnis zusammenfassend kann gesagt werden: Die Einbeziehung der Variablen "Geschlecht" verändert die betrachtete Kausalstruktur zwar, doch schränkt diese Geschlechtsdifferenz den Geltungsbereich der Hypothese von der entscheidenden Vermittlung durch die Kurszugehörigkeit nicht ein. Es liegt nur eine Interaktion derart vor, dass sich bei den Jungen die Eltern-Unterstützung nicht auf die Lehrer-Unterstützung auswirkt.

4. ZUSAMMENFASSUNG UND DISKUSSION DER ERGEBNISSE

4.1. Ergebnisse

Ein Ziel der Untersuchung war es zu testen, ob die Förderstufe ihr Ziel "Fördern statt Auslesen", bezogen auf die Anstrengungsbereitschaft, erreichen kann, wenn ihr Kurssystem relativ stabil quasi-streaming organisiert ist.

Der Test führte in zwei voneinander unabhängigen Stichproben, einer Gruppe von Jungen und einer Gruppe von Mädchen, zum gleichen Ergebnis: Die Korrelationen zwischen den Kurszugehörigkeiten der Schüler und ihrer Anstrengungsbereitschaft sind am Ende der Förderstufenzeit so (in Höhe von 0,46 bis 0,56), dass sie auch bei der simultanen Kontrolle aller miterhobenen Variablen (soziale Herkunft, Testintelligenz, Eltern-Unterstützung, Lehrer-Unterstützung), deren kausale Reihung mitberücksichtigt, nicht verschwinden.

Im Zusammenhang mit weiteren Resultaten der Kausalanalyse deutet dies, bezogen auf die Anstrengungsbereitschaft, auf eine Unverträglichkeit von Selektion und Förderung hin: Es konnte in beiden Geschlechtergruppen die Hypothese nicht widerlegt werden, dass der Einfluss der sozialen Herkunft auf die Anstrengungsbereitschaft (im Rahmen des vorgegebenen Kausalmodells) entscheidend durch die Kurszugehörigkeit vermittelt wird. Im Hinblick auf den beabsichtigten Ausgleich kursniveauspezifischer Anstrengungsbereitschaften war die betrachtete Förderstufen-Schule offensichtlich nicht erfolgreich.

Das vorgenannte kausalanalytische Ergebnis widerspricht auch der Vermittlungshypothese von BORCHERT & MASENDORF, die besagt, die Unterstützung durch die Mutter und durch die Lehrer vermittelten den Einfluss der sozialen Herkunft eines Schülers auf seine Anstrengungsbereitschaft.

4.2. Zur Gültigkeit und Bedeutung der Ergebnisse

Nach den Ergebnissen spielen bei der Förderung der Anstrengungsbereitschaft die Lehrer eine wenn auch nicht entscheidende, so doch mitverantwortliche Rolle: Sie unterstützen die C-Kurs-Schüler (unabhängig von deren sozialer Herkunft und Testintelligenz) weniger als die höher eingestuften Schüler. Dieser Befund widerspricht zugleich zwei verschiedenen Auffassungen von "Förderung": Es ist nicht der Fall, dass den C-Kurs-Schülern stärkere Unterstützung zuteil wird als den höher eingestuften Schülern, und es ist auch nicht der Fall, dass alle Schüler wenigstens gleiche Unterstützung erfahren.

Die Tatsache, dass wir das Unterstützungsverhalten der Lehrer indirekt über die Eindrücke der Schüler erfassten, mag man zum Anlass nehmen darauf hinzuweisen, dass die Aussagen der Schüler auch Ergebnisse von Abwehrmechanismen (z.B. Projektion, Verleugnung) sein könnten: Aufgrund seiner schlechten Leistungsposition neige der C-Kurs-Schüler dazu, sich vom Lehrer nicht akzeptiert und unterstützt zu fühlen, auch wenn dies aktuell gar nicht zutreffen muss. Zur eigenen nachträglichen Rechtfertigung seiner niedrigen Leistungsposition nehme der Schüler auch hinter Wohlwollen versteckte Ablehnung wahr, tatsächlich gegebene Zuwendung und Unterstützung werde aufgrund seiner inferioren Lage "weg-interpretiert".

Diese Schüler-Reaktion ist aber nur verständlich, wenn man zusätzlich unterstellt, dass die Schüler ihre niedrige Leistungsposition selbstwertbezogen interpretieren. Mit den hierarchisch angeordneten Leistungsgruppen sind in unserem Schulsystem bestimmte soziale Wertschätzungen verknüpft. In diese Bewertungen fliessen Voreingenommenheiten in Bezug darauf, "was man einem zutraut", mit ein. Die mit seiner Leistungsposition verknüpften sozialen Wertschätzungen und Leistungserwartungen - oft unausgesprochen und subtil kommuni-

ziert - nimmt der Schüler, wie einige Ergebnisse belegen (s. z.B. BROPHY & GOOD, 1970), sehr deutlich wahr. Wenn man aber mit SCHWARZER & ROYL (1976, S. 548) davon ausgeht, dass der Leistung in der Schule als wesentlichem Mittel zur Erlangung von sozialer Anerkennung und emotionaler Zuwendung besonderer Wert beigemessen wird, dann fehlt in einer schulischen Situation, in der der Schüler durchgängig spürt, wie wenig Lehrer (und auch Eltern) von seiner Leistungsfähigkeit überzeugt sind, auch die emotionale Grundlage dafür, dass er sich Leistungen zutraut.

Mit anderen Worten, wie wir die Eindrücke der Schüler zum Lehrerverhalten auch auffassen, als Reflektionen des Lehrerverhaltens oder als Ergebnisse von Abwehrmechanismen: Die in der stabilen Kurseinstufung sich ausdrückende Selektion ist, so erleben das die C-Kurs-Schüler, mit ihrer Förderung nicht vereinbar.

Machen diese Erklärungen möglicherweise auch nur verständlich, warum gerade in einem integrierten Schulsystem Schüler unterer Leistungsgruppen Motivationsprobleme haben, so darf dabei eines nicht übersehen werden: Die Rede von der unzureichenden Förderung einer Schülergruppe in der Förderstufe bezieht sich auf das selbstgesteckte Ziel dieser Schule und sagt über die Ueberlegenheit des segregierten Schulsystems noch garnichts aus.

Einschränkungen der vorliegenden Arbeit liegen aussserdem in der regional begrenzten Erhebungssituation. Wir betrachteten einen Schülerjahrgang in nur einer Förderstufen-Schule. Zur Widerlegung einer als allgemeingültig gedachten Hypothese war das hinreichend, zum Zwecke der Generalisierung der Ergebnisse aber weniger. Trotzdem stellt sich die Frage, ob die Befunde z.B. auch für andere Förderstufen-Schulen gelten können. Auf andere quasi-streaming organisierte Förderstufen-Schulen in der Stadt Wetzlar erscheint das Ergebnis übertragbar, zumal wir aufgrund der erfolgreichen Kreuzvalidierung des Testergeb-

nisses annehmen dürfen, dass wir in unserer Stichprobe nicht einmalige Verhältnisse vorgefunden haben. Ungewiss ist auch, ob die Ergebnisse für entsprechende Förderstufen ausserhalb der Stadt Wetzlar Gültigkeit beanspruchen dürfen. Denn die Förderstufen-Situation in Wetzlar ist, abgesehen vielleicht von derjenigen in Hanau und Kassel, durch einige Besonderheiten gekennzeichnet. Zwei davon seien kurz genannt: In Wetzlar ist die Förderstufe obligatorisch und ihr Lehrerkollegium rekrutiert sich möglicherweise nur zum geringeren Teil aus engagierten Versuchsschullehrern. Viel grösser als bei den Angebots-Förderstufen ist hier vermutlich die Zahl der Lehrer, die als ehemalige Hauptschul-, Realschul- und Gymnasial-Lehrer eher notgedrungen in der neuen Schulform mitzuarbeiten hatten.

Die Ungewissheit in Bezug auf überregionale Uebertragbarkeit der Ergebnisse schränkt aber deren praktische Bedeutung nicht notwendigerweise ein. Eher wäre den Ergebnissen eine vergleichsweise geringere ökologische Validität (im Sinne BRUNSWIKs, 1955) zuzuschreiben, wären sie nicht in Wetzlar gewonnen. Denn in Wetzlar wird ja die Förderstufe unter Bedingungen erprobt, unter denen sie später auch ausserhalb Wetzlars angewendet werden soll: Die Förderstufe als schulischer Pflichtfall.

Wichtiger als das Problem der regionalen Generalisierbarkeit - wenn auch in unserem Untersuchungszusammenhang des Testens allgemeiner Hypothesen zunächst sekundär - erscheint die Frage, auf welchen pädagogischen Kontext das Ergebnis ausgedehnt werden kann, z.B. ob es ähnlich auch innerhalb des segregierten Schulsystems - mit den Schularten als den "Leistungsgruppen" - gilt. Es ist durchaus denkbar, dass wir in unserer speziellen Untersuchungssituation eine allgemeine Hypothese falsifiziert haben, ohne zunächst - beim Experiment ist das nicht anders - zu wissen, was genau unter "ähnlichen" Bedingungen zu verstehen ist. Jede Schlussfolgerung gilt daher nur begrenzt. Dennoch kann sie von Interesse sein.

Bevor zu Schlussfolgerungen übergegangen wird, ist - um eventuellen Missverständnissen vorzubeugen - der hier gemeinte Begriff "Schlussfolgerung" zu explizieren. Es erscheint wichtig, dabei mindestens drei Punkte zu beachten:

1. Eine Kausal- bzw. Nullhypothese, als "nicht verworfen" im Modellvergleich beurteilt, ist hier nicht als äquivalent mit "als richtig festgestellt" zu werten. Aber wir gelangen zu dem praktischen Schluss, dass wir, wären wir gezwungen, pädagogisch oder bildungspolitisch zu handeln, so handeln müssten, als ob die Nullhypothese richtig wäre. Das jedoch ist nicht dasselbe wie die Feststellung: "Die Nullhypothese ist richtig". Im Sinne von: "Die Nullhypothese ist richtig" sind die Schlussfolgerungen nicht zu verstehen.

   "Nicht verworfen" kann in diesem Untersuchungszusammenhang auch nicht heissen, dass das "optimale" oder das am besten auf die Daten passende Modell "gefunden" worden ist. Die hier an Entscheidungsmodelle zu richtenden Ansprüche zielten nicht auf das Finden eindeutig optimaler Lösungen ab.

2. Die Schlussfolgerungen ergeben sich aus der Erwünschtheit bzw. Nicht-Erwünschtheit pädagogischer Resultate. Ausgangspunkt ist die Feststellung, dass man mit einem relativ stabilen quasi-streaming organisierten Kurssystem die gewünschte kompensatorische Förderung der Anstrengungsbereitschaft nicht erreicht. Dem Untersuchungsergebnis entsprechend ist über einige Handlungsmöglichkeiten zu informieren, welche der Förderstufe helfen könnten, die selbstgesteckten Ziele besser zu erreichen als bisher. Mit dem Urterschied zwischen integriertem und segregiertem Schulsystem oder gar mit einem Plädoyer für Abschaffung der Förderstufe haben diese Schlussfolgerungen absichtlich nichts zu tun, eher damit, inwieweit für eine kompensatorische Förderung der Anstrengungsbereitschaft in bestimmten Aspekten das ganze Bildungssystem reformbedürftig ist.

3. Wenn aufgrund der Ergebnisse z.B. gefolgert wird, dass nicht homogen (leistungs-)differenziert werden sollte, so geschieht dies nicht, weil in der Untersuchung eine Ueberlegenheit der heterogenen über die homogene Leistungsgruppierung festgestellt worden wäre. Die Schlussfolgerungen sind hier Erklärungshypothesen, welche nur deshalb den Status von Schlussfolgerungen erhalten, weil nach einer erfolgreichen Prognose zum Zwecke besserer präventiver Intervention folgerichtig an die Erklärungsversuche bzw. Zusatzannahmen angeknüpft wird, die zur Herleitung eben dieser Prognose notwendig waren.

   Wesentlich war die Zusatzannahme, dass die relativ stabile quasi-streaming organisierte Kurseinstufung eine für den C-Kurs-Schüler verhängnisvolle Kombination zweier Bedingungen enthält, nämlich "gruppenbezogene Leistungsbeurteilung" und "Ueberbewertung der Kursfächer", d.h. es dürfte eine Erklärung für die beobachtete geringere Anstrengungsbereitschaft der C-Kurs-Schüler darin liegen, dass es dem im Kurssystem niedrig eingestuften Schüler ziemlich schwerfällt, seine Leistungsposition nur im Hinblick auf die Leistungsverhältnisse in der eigenen Gruppe, also der C-Kurs-Gruppe, zu beurteilen. Dies wiederum dürfte damit zusammenhängen, dass die Förderstufe (als Institution) den Schulerfolg hauptsächlich an den Leistungen in den Kursfächern abliest.

   Sicherlich ist mit diesen Erklärungen noch nicht genau spezifiziert, in welchem Ausmass die beiden Bedingungen "gruppenbezogene Leistungsbeurteilung" und "Ueberbewertung der Kursfächer" im einzelnen zum Resultat beigetragen haben. Doch war es in unserem Untersuchungszusammenhang auch nicht notwendig, operational so vorzugehen, dass zwischen den hier erwogenen Bedingungen eindeutig hätte entschieden werden können. Denn zum einen war zu erkennen, dass sie auch im Kurssystem der Förderstufe nicht isoliert voneinander betrachtet werden konnten. Es wäre auch wenig sinnvoll, sich

über eventuelle pädagogische Konsequenzen der einen Bedingung Gedanken zu machen, wenn man mögliche Wechselwirkungen zwischen beiden nicht mitberücksichtigt.

Andererseits erlaubte aber gerade die gemeinsame Betrachtung dieser beiden als wirksam vermuteten Bedingungen die theoretisch begründete Ableitung erfolgreicher Prognosen. Genaugenommen wurde hier daher der Ausdruck "entscheidende Vermittlung durch die Kurszugehörigkeit" im Rahmen einer bislang praktischen Situation gebraucht, in der es zunächst nur darauf ankam, etwas herauszufinden. Es sollte herausgefunden werden, ob sich eine mit Hilfe der Zusatzannahmen abgeleitete alternative Vorhersage im Modellvergleich bewährt. In diesem Sinne konnte eine eindeutige Entscheidung getroffen werden: Die Hypothese von der entscheidenden Vermittlung durch die Kurszugehörigkeit, gleichgültig, wie komplex auch immer die Kurszugehörigkeit aus den erwogenen Bedingungen "zusammengesetzt" sein mag, war im Modellvergleich nicht zu widerlegen.

Da die Daten nicht ausreichen, alle in die Herleitung der Kausalhypothese eingehenden Zusatzannahmen zu prüfen, gilt auch hier, dass wir, wären wir gezwungen zu handeln, dem Ausgang des Tests entsprechend zunächst so handeln müssten, als ob diese Zusatzannahmen richtig wären. Es muss nicht noch besonders betont werden, dass pädagogisch-praktisches Wissen immer nur in vorläufigen und versuchsweisen Lösungsvorschlägen besteht und daher prinzipiell die Möglichkeit einschliesst, dass es sich als irrtümlich herausstellen wird. Bei dieser Art des Vorgehens ist zu berücksichtigen, dass die Sachzwänge das Treffen von Entscheidungen ständig erfordern und dass diese Entscheidungen auch getroffen werden, wenn noch kein hyperkritisch geforderter Eindeutigkeitsbeweis, noch kein vollständiges oder optimales Problemlösungsverständnis vorliegt.

4.3. Pädagogische Folgerungen

Unter der Voraussetzung, dass die Ergebnisse in wiederholten Untersuchungen, auch unter Einbezug späterer Untersuchungszeitpunkte (z.B. Untersuchung derselben Schüler auch in der Sekundarstufe I einer integrierten Gesamtschule) nicht viel anders ausfallen, lässt sich allgemein formuliert, folgendes ableiten: Unter der Bedingung eines relativ stabilen quasi-streaming organisierten Kurssystems kann die Förderstufe kompensatorische Förderung der Anstrengungsbereitschaft nicht erreichen.

Es liegt eine eher resignative Lösung dieses unerwünschten Sachverhalts nahe: Wenn man kompensatorische Förderung unter dieser Bedingung nicht erreichen kann, sollte man kompensatorische Förderung auch nicht betreiben. Geben wir jedoch den Anspruch der Förderstufe nicht so ohne weiteres preis, dass Anstrengungsbereitschaft kompensatorisch gefördert werden sollte, und halten wir an der im Anspruch enthaltenen optimistischen Auffassung vorläufig fest, dass kompensatorische Förderung der Anstrengungsbereitschaft von der Förderstufe auch betrieben werden kann: Was ist dann zu tun?

Es ist ein Ergebnis dieser Untersuchung, dass in einem relativ stabilen quasi-streaming organisierten Kurssystem offenbar auch bei Kontrolle des Lehrerverhaltens direkte Effekte ("direkt" bezogen auf das vorgegebene Kausalmodell) der Kurszugehörigkeit auf die Anstrengungsbereitschaft der Schüler bestehen. Die hier beobachtete Unverträglichkeit von Selektion und Förderung kann vom Lehrer anscheinend nicht individuell in seinem Unterricht "aufgehoben" werden. Darum legen die Ergebnisse, nimmt man das Ziel der kompensatorischen, individuellen Förderung ernst, zunächst einmal den Schluss nahe, Fördermassnahmen nicht allein über die Person des Lehrers zu versuchen.

Es leuchtet zwar unmittelbar ein, dass eine resignierte Lehrerschaft für die Förderstufenentwicklung ein besonderes Verhängnis wäre. Wenn man aber nur dafür plädiert, zunächst reformfreudigere, engagiertere Lehrer einzusetzen (z.B. ROTH, 1975) oder erst einmal die die Lehrer belastende Verwaltungsarbeit abzubauen, um zu verwertbaren Ergebnissen zu kommen, dann ist das, nimmt man unsere Ergebnisse beim Wort, offenbar zu wenig. Man muss eher damit rechnen, dass die beste sozialpsychologische Lehrerausbildung bzw. der engagierteste Lehrereinsatz die mit dem Kurssystem zusammenhängenden Motivationsprobleme der C-Kurs-Schüler bestenfalls mildern, doch keineswegs lösen würde. So notwendig einstellungs- und verhaltensbezogene Trainingsprogramme für Lehrer auch sind, schulische Fördermassnahmen wären möglicherweise nur ein Kurieren an Symptomen, würden sie nur am Lehrer ansetzen und nicht auch institutionelle Bedingungen der Schule miteinbeziehen.

Wenigstens die Richtung soll hier angedeutet werden, in der eine solche Veränderung gehen könnte. Wir knüpfen dazu an die Zusatzannahmen an, mit deren Hilfe es möglich war, eine im Modellvergleich erfolgreiche Kausalhypothese abzuleiten. In die von diesen Zusatzannahmen, von diesen Anknüpfungspunkten ausgehenden Lösungsvorschläge fliessen natürlich mehr oder weniger persönliche Ueberzeugungen und Meinungen mit ein.

Zu denken wäre zunächst einmal an einen Verzicht auf die starre vertikale Differenzierung nach Leistung in den Kursfächern. Durch die Zuweisung in den C-Kurs einer Förderstufe versprachen sich die Schulreformer eine Förderung des Schülers. Möglicherweise wurde dabei aber übersehen, dass diese Zuweisungen je nach der subjektiven Definition der schulischen Situation durch den betroffenen Schüler anders wahrgenommen und erlebt werden können und damit auch anders als beabsichtigt wirken. Unabhängig davon, was beabsichtigt ist: Das Kurssystem scheint dadurch gekennzeichnet, dass seine unteren Leistungsgruppen

eine schlechtere soziale Wertschätzung erfahren in dem Sinne, dass nicht die Zugehörigkeit, sondern letztlich das Verlassen dieser Leistungsgruppe "nach oben" als erwünscht gilt. Wenn der Masstab für die Einschätzung der eigenen Leistung das gesamte Kurssystem ist, dann dürfte die Zuweisung zu einer niedrig bewerteten Leistungsgruppe die Anstrengungsbereitschaft um so eher abbauen, je unabänderlicher diese Zuweisung bestimmte Aufstiegsmöglichkeiten ausschliesst.

Durch eine stabile äussere Differenzierung wird der C-Kurs-Schüler möglicherweise fortwährend darauf hingewiesen, um wieviel er schlechter ist als andere Schüler. Viele Hoffnungen knüpfen sich daher an die sog. "innere Differenzierung" des Unterrichts (s. dazu KLAFKI & STOECKER, 1976). Solange man nach Leistungen differenziert, sollte man das wenn irgend möglich nur innerhalb der Lerngruppe versuchen. Denn der Tendenz nach scheint es, zieht man andere Untersuchungen zu Rate, eher so zu sein, dass leistungsschwache Schüler in heterogen zusammengesetzten Lerngruppen bessere Schulerfolge - und es ist auch anzunehmen, bessere Anstrengungsbereitschaften - erreichen als in leistungshomogenen Niveaugruppen (z.B. RANG & SCHULZ, 1969; PRELL, SCHIEFELE & ULICH, 1972).

Um die Förderung leistungsschwacher Schüler kann man sich zusätzlich bemühen, wenn man bei der Lernorganisation - dort, wo es möglich und sinnvoll ist - vom Gesichtspunkt des "zielerreichenden Lernens" (BLOOM, 1973; CARROLL, 1973) ausgeht. Man könnte so zu gewährleisten versuchen, dass individuelle Lernanstrengungen nicht vergeblich bleiben, sondern ein Leistungsziel in wiederholten Versuchen auch erreicht werden kann. Dies kann geschehen, ohne das Fortkommen der schneller lernenden Kinder zu behindern (s. dazu TREIBER et al., 1976). Der Erfolg, den der Schüler erlebt, wenn er ein festgesetztes Leistungsziel erreicht, müsste dazu beitragen, dass er lernt, sich wieder etwas zuzutrauen.

Das bedeutet auch, dass dem Lehrer zumindest in der Förderstufenzeit institutionell ein grösserer Handlungsspielraum zugestanden werden müsste als bisher. Denn er müsste nicht nur lehren, sondern auch helfen und beraten. Er müsste in die Lage versetzt werden, dem Schüler das Gefühl zu vermitteln, dass selbständige Leistungsbemühungen auch wirklich prämiiert werden, ob nun die Leistungsfortschritte weiter reichen oder weniger weit: Der Schüler sollte sich ohne Disqualifizierung seiner Leistungsbemühungen akzeptiert und unterstützt fühlen können.

Anzustreben wären deshalb auch konstante Kleingruppen, die in mehreren Unterrichtsfächern zusammenbleiben. Der Lehrer wäre so eher imstande, in seinen Bekräftigungen stärker auf die subjektive Bedeutung der Leistungshandlung für die Schüler einzugehen, z.B. auf die Kausalattributionen, die die Schüler über das Zustandekommen ihrer Leistungsresultate haben, was besonders den leistungsschwachen Schülern helfen könnte, ein die Anstrengungsbereitschaft förderndes Selbstwertgefühl aufzubauen. In grossen Schulklassen fällt es Lehrern offenbar schwer, den Schwierigkeitsgrad von Aufgaben für ihre Schüler richtig einzuschätzen (vgl. z.B. FLUEGGE, 1954). In der Förderstufe wird dieses Problem durch den mit der Kern-Kurs-Organisation verbundenen häufigen Lehrerwechsel möglicherweise noch verschärft. Wenn man aber die Schwierigkeiten der Aufgaben für die Schüler nicht kennt, kann Lob und Ermunterung durch den Lehrer vor allem bei leistungsschwachen Schülern nicht in wirksamer Weise angewendet werden.

Zielerreichendes Lernen ist nicht ganz unproblematisch und hat sicherlich auch seine Grenzen. So kann z.B. "Erfolg für alle" kaum vermittelt werden, wenn Leistungsergebnisse in nur zwei Fächern (den Kursfächern) fast ausschliesslich zum sozialen Vergleich herangezogen werden. Selbst wenn die Lehrer gegenüber den Schülern kaum sozial wertende Leistungsvergleiche

anstellen würden, dann dürfte bereits das Ausmass vollauf genügen, in dem die Schüler dies von sich aus tun, d.h. auch unter dem Aspekt zielerreichenden Lernens sind bei aller guten Absicht der Lehrer disqualifizierende Schematisierungseffekte nicht auszuschliessen. Es wäre auch zu einfach, würde man nur die Organisation schulischen Lernens als zu ändernde Variable ansehen. Die Schule bleibt bei allen derartigen Organisationen des Lernens in dem Dilemma, gleichzeitig Grundqualifikationen (Allgemeinbildung) entwickeln und auch besondere Neigungen und Fähigkeiten fördern zu müssen.

Hält man sich die massgebliche Bedeutung der Kursfächer für den Schulabschluss und für die Berufslaufbahn vor Augen, dann muss man daran zweifeln, ob die Unterschiede in der Anstrengungsbereitschaft durch allein auf die Schule bezogene Reformmassnahmen je werden ausgeglichen werden können, dies vor allem dann, wenn man analog wie bei der sog. "Defizithypothese" der Intelligenz unterstellt, dass überhaupt ein Mangel an Anstrengungsbereitschaft auszugleichen sei. Fasst man mangelnde Anstrengungsbereitschaft als (bisher) nicht gelungene Anpassung an institutionell definierte Anforderungen auf, dann muss man eher davon ausgehen, dass die von uns gemessene Anstrengungsbereitschaft nicht allgemeingültig, sondern fachspezifisch ist, d.h. der benutzte Test verzeichnet nur, wie die Schüler auf die von der Schule hier und heute für wichtig gehaltenen Leistungsanforderungen reagieren, und es ist nicht anzunehmen, dass die Schule mit ihrer ungleichen Fächergewichtung eine für alle Schüler gleiche Anregungsbedingung für Anstrengungsbereitschaft darstellt. So reagieren gute und schlechte Schüler angesichts derselben Leistungsanforderungen unterschiedlich: Anstrengungsbereitschaft lässt sich nur schwer entwickeln in Leistungsbereichen, in denen permanent Misserfolge erlebt werden. In welchen Leistungsbereichen die Anstrengungsbereitschaft schlechter Schüler angeregt würde, erfährt man ja nicht, wenn sie sich einem Test unterziehen, der vorwiegend Leistungssitua-

tionen zum Gegenstand hat, in denen sie Misserfolge erleben.

Auf der Grundlage dieser Ueberlegungen lässt sich aus den vorliegenden Ergebnissen auch der Hinweis entnehmen, dass man auf kompensatorisches Lehrerverhalten unter Umständen verzichten könnte, wenn die institutionelle Anforderungsstruktur der Schule pädagogisch angemessen geändert würde. Uns scheint, die übermässig hohe Wertschätzung sprachlicher und mathematischer, also intellektuell-abstrahierender Leistungen innerhalb der quasi-streaming organisierten Förderstufe macht zugleich ihr Wahlangebot zunichte. Kern- bzw. Wahlfächer können vom Schüler wohl schwerlich ernstgenommen werden, wenn die Schule mit ihrer einseitigen Ausrichtung auf ein gymnasial-akademisches Förderungsziel Leistungen in nur wenigen Pflichtfächern auszeichnet vor allen anderen Leistungen (u.a. dadurch, dass Niveaukurse in diesen Pflichtfächern eingerichtet werden). Es liegt nahe, dass so die Chancen reduziert werden, die eigenen Fähigkeiten zu erproben und Anstrengungsbereitschaft fördernde Schulerfolge zu erleben. Da höhere Lebenschancen primär durch gute Schulleistungen erworben werden, bedeutet eine Höherbewertung sprachlicher und mathematischer Leistungen eine Benachteiligung für diejenigen Schüler, die ihre Stärken auf ganz anderen Gebieten haben, also z.B. für die gestalterisch-, handwerklich-, technisch-, sozialhelferisch-begabten Schüler, um nur einige zu nennen. Im Interesse dieser Schüler wäre eine Chancenvermehrung bzw. Korrektur der Fächerbewertungen am Platze. So sollten Leistungserfolge in den Kern- bzw. Wahlfächern ebenfalls mit Prestige(zuwachs) belohnt werden, allgemein: Wenn die Schüler differenzierende Leistungsangebote erhalten, die nicht von vornherein in höhere und niedere für leistungsstarke und leistungsschwache Schüler unterschieden werden, dann ist zu erwarten, dass sich Schüler mit verschiedenen Fähigkeiten in der Schule eine annähernd gleiche Anregung ihrer Anstrengungsbereitschaft bietet.

All dies ist natürlich leichter gesagt als getan. Man sollte nicht davon ausgehen, dass Schule den gesellschaftlichen Zwang zur Auslese nach Belieben mildern kann. Da allgemeine gesellschaftliche Erwartungen und Selbstverständlichkeiten stets auch in die Schule durchschlagen, setzt eine in der Schule wirksame Veränderung z.B. von Fächerbewertungen auch eine gewisse Veränderung der gegenwärtig in unserer Gesellschaft geltenden Wertehierarchie voraus, mit anderen Worten: erst müssten sich neue Leitbilder beruflichen Prestigestrebens entwickeln, damit vom differenzierenden Leistungsangebot nicht nur derjenige Schüler profitiert, der das Lernen bereits gelernt hat, sondern auch der, der seine Lernvoraussetzungen, seine Anstrengungsbereitschaft noch verbessern soll.

5. ZUSAMMENFASSUNG

Ein Ziel der Untersuchung war es zu testen, ob die Förderstufe ihr Ziel "Fördern statt Auslesen", bezogen auf die Anstrengungsbereitschaft, erreichen kann, wenn ihr Kurssystem relativ stabil quasi-streaming organisiert ist.

Die Daten dazu wurden an einer Stichprobe von 144 Schülern einer Gesamtschule in der Stadt Wetzlar erhoben.

Analysen von Strukturgleichungsmodellen erbrachten in zwei voneinander unabhängigen Stichproben, einer Gruppe von Jungen und einer Gruppe von Mädchen, das gleiche Ergebnis: Es konnte die Hypothese nicht widerlegt werden, dass der Einfluss der sozialen Herkunft auf die Anstrengungsbereitschaft (im Rahmen des vorgegebenen Kausalmodells) entscheidend durch die Kurszugehörigkeit vermittelt wird. Im Hinblick auf den beabsichtigten Ausgleich kursniveauspezifischer Anstrengungsbereitschaften war die betrachtete Förderstufen-Schule offensichtlich nicht erfolgreich.

Das kausalanalytische Ergebnis widerspricht auch der Vermittlungshypothese von BORCHERT & MASENDORF, die besagt, die Unterstützung durch die Mutter und durch die Lehrer vermittelten den Einfluss der sozialen Herkunft eines Schülers auf seine Anstrengungsbereitschaft.

Die Ergebnisse deuten darauf hin, dass auch der engagierteste Einsatz von Seiten der Lehrer die Motivationsprobleme der im Kurssystem der Förderstufe niedrig eingestuften Schüler nicht lösen würde. Denkbar ist darum auch, dass den Problemen einer Neudefinition der schulischen Leistung besondere Aufmerksamkeit zu widmen ist.

6. LITERATURVERZEICHNIS

ALLEN, M.P.: Construction of Composite Measures by the Canonical-Factor-Regression Method. In: COSTNER, H.L. (ed.): Sociological Methodology 1973/74. San Francisco: Jossey-Bass, 1974, 51-78

ANDERSON, G.J.: Effects of Classroom Social Climate on Individual Learning. In: Amer. educ. Res. J., 1970, 7, 135-152

ANDO, A. & FISHER, F.M.: Near-Decomposability, Partition and Aggregation, and the Relevance of Stability Discussions. In: Internat. Econ. Rev., 1963, 4, 53-67

AUSUBEL, D.P.: Theory and Problems of Child Development. New York: Grune & Straton, 1958

BACH, W.: Der Aufbau der Förderstufe in Hessen - Modell Wetzlar. In: BACH, W. (Hg.): Der Auftrag der Schulpsychologie für die Schule von morgen. Weinheim: Beltz, 1972, 143-152

BAUER, A.: Ein Verfahren zur Messung des für das Bildungsverhalten relevanten sozialen Status (BRSS). Frankfurt/M.: Unveröffentlichtes Manual aus dem Deutschen Institut für Internationale Pädagogische Forschung, 1972

BECKER, H.S.: Social-Class Variations in the Teacher-Pupil Relationship. In: BELL, R.R. & STUB, H.R. (eds.): The Sociology of Education. Homewood (Ill.), 1968, 155-166

BERNHARDT, M., BOETTIGER, M., van HOLST, H.D., KACZENSKI, G. & WEIGELT, K.-G.: Soziales Lernen in der Gesamtschule. München: Juventa, 1974

BLALOCK, H.M., Jr.: The Measurement Problem: A Gap between the Languages of Theory and Research. In: BLALOCK, H.M., Jr. & BLALOCK, A.B. (eds.): Methodology in Social Research. New York: McGraw-Hill, 1968, 5-27 (a)

BLALOCK, H.M., Jr.: Theory Building and Causal Inferences. In: BLALOCK, H.M., Jr. & BLALOCK, A.B. (eds.): Methodology in Social Research. New York: McGraw-Hill, 1968, 155-198 (b)

BLALOCK, H.M., Jr.: Theory Construction. From Verbal to Mathematical Formulations. Englewood Cliffs (N.J.): Prentice-Hall, 1969

BLOOM, B.S.: Individuelle Unterschiede in der Schulleistung: ein überholtes Problem? In: EDELSTEIN, W. & HOPF, D. (Hg.): Bedingungen des Bildungsprozesses. Stuttgart: Klett, 1973, 251-270

BORCHERT, J.: Empirische Untersuchungen über Beziehungen zwischen Erziehungsverhalten von Lehrern und Müttern und seiner Rezeption durch die betroffenen Schüler im Hinblick auf Schulangst, Anstrengungsvermeidung und soziale Beliebtheit. Dortmund: Dissertation an der Pädagogischen Hochschule Ruhr, 1974

BORCHERT, J. & MASENDORF, F.: Zum Bedingungsverhältnis von Sozialstatus, Erziehungsstil der Mutter und des Lehrers und affektiven Schülervariablen. In: Psychol. in Erzieh. u. Unterricht, 1975, 22, 137-147

BROPHY, J.E. & GOOD, T.L.: Teacher's Communication of Differential Expectations for Children's Classroom Performance: Some Behavioral Data. In: J. educ. Psychol., 1970, 61, 365-374

BRUNSWIK, E.: Representative Design and Probabilistic Theory in a Functional Psychology. In: Psychol. Rev., 1955, 62, 193-235

BRUSTEN, M. & HURRELMANN, K.: Abweichendes Verhalten in der Schule. München: Juventa, 1973

CAMPBELL, D.T. & STANLEY, J.C.: Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963

CARROLL, J.B.: Lernerfolg für alle. In: HOFER, M. & WEINERT, F.E. (Hg.): Pädagogische Psychologie. Grundlagentexte 2: Lernen und Instruktion (Funkkolleg). Frankfurt/M.: Fischer, 1973, 124-135

COLEMAN, J.S.: The Mathematical Study of Change. In: BLALOCK, H.M., Jr. & BLALOCK, A.B. (eds.): Methodology in Social Research. New York: McGraw-Hill, 1968, 428-478

DAHRENDORF, R.: Gesellschaft und Demokratie in Deutschland. München: Piper, 1968

DAVIDSON, H.H. & LANG, G.: Children's Perceptions of their Teacher's Feelings toward them Related to Self-Perception, School Achievement and Behavior. In: J. Exper. Educ., 1960, 29, 107-118

DUNCAN, O.D., FEATHERMAN, D.L. & DUNCAN, B.: Socioeconomic Background and Achievement. New York & London: Seminar Press, 1972

FEND, H., KNOERZER, W., NAGL, W., SPECHT, W. & VAETH-SZUSDZIARA, R.: Zwischenbilanz eines Schulversuchs. Ergebnisse einer empirischen Vergleichsuntersuchung zwischen Gesamtschulen und dreigliedrigem Schulsystem. In: Gesamtschule, 1975 (Mai), 6-16

FERDINAND, W. & KIWITZ, H.: Ueber die Häufigkeitsverteilung der Zeugnisnoten 1 bis 6. In: INGENKAMP, K. (Hg.): Die Fragwürdigkeit der Zensurengebung. Weinheim: Beltz, 1971, 178-185

FISHER, F.M.: The Identification Problem in Econometrics. New York: McGraw-Hill, 1966

FLUEGGE, I.: Die Einschätzung der Schwierigkeiten von Aufgaben durch den Lehrer. Frankfurt/M.: Sonderheft der Mitteilungen und Nachrichten der Hochschule für Internationale Pädagogische Forschung, 1954

GEISSLER, E.E., KRENZER, R.P. & RANG, A.: Fördern und Auslesen. Frankfurt/M.: Diesterweg, 1967

GOLDBERG, M.L., PASSOW, H. & JUSTMANN, J.: Auswirkungen der Niveaugruppenbildung. In: RANG, A. & SCHULZ, W. (Hg.): Die differenzierte Gesamtschule. München: Piper, 1969, 37-60

GOLDBERGER, A.S.: Econometric Theory. New York: Wiley, 1964

GRESSER-SPITZMUELLER, R.: Lehrerurteil und Bildungschancen. Weinheim: Beltz, 1973

HARMAN, H.H.: Modern Factor Analysis. Chicago: The University of Chicago Press, 1967

HECKHAUSEN, H.: Förderung der Lernmotivierung und der intellektuellen Tüchtigkeiten. In: ROTH, H. (Hg.): Begabung und Lernen. Stuttgart: Klett, 1968, 193-228

HECKHAUSEN, H.: Die Interaktion der Sozialisationsvariablen in der Genese des Leistungsmotivs. In: GRAUMANN, C.F. (Hg.): Sozialpsychologie (Handbuch der Psychologie, 7. Band). Göttingen: Hogrefe, 1972, 955-1019

HEINRICH, H.C.: Skalen zur Erfassung von Formen der Bekräftigung in der Erziehung durch Lehrer. In: Zt. exp. angew. Psychol., 1974, 21, 530-545

HEISE, D.R. & BOHRNSTEDT, G.W.: Validity, Invalidity, and Reliability. In: BORGATTA, E.F. & BOHRNSTEDT, G.W. (eds.): Sociological Methodology 1970. San Francisco: Jossey-Bass, 1970, 104-129

HEPTNER, B.E.: Schule der Zukunft. In: Frankfurter Allgemeine Zeitung (FAZ) vom 10.6.1976

HERRENKOHL, R.C.: Factor-Analytic and Criterion Study of Achievement Orientation. In: J. educ. Psychol., 1972, 63, 314-326

HERRMANN, T., SCHWITAJEWSKI, E. & AHRENS, H.J.: Untersuchungen zum elterlichen Erziehungsstil: Strenge und Unterstützung. In: Arch. ges. Psychol., 1968, 120, 74-105

HIELSCHER, H.: Unterrichtsdifferenzierung und Lehrerverhalten. In: Westermanns Pädag. Beitr., 1970, 22, 621-629

HOFER, M.: Die Schülerpersönlichkeit im Urteil des Lehrers. Weinheim: Beltz, 1969

HOPP, A.-D. & LIENERT, G.A.: Eine Verteilungsanalyse von Gymnasialzensuren. In: Schule u. Psychologie, 1965, 12, 139-150

HUESKE, H.: Soziales Lernen als Psychohygiene. Vermittlung emotional-korrigierender Erfahrungen im Englischunterricht der Gesamtschule. In: KEIM, W. (Hg.): Gesamtschule. Bilanz ihrer Praxis. Hamburg: Hoffmann & Campe, 1973, 345-359

HURRELMANN, K.: Unterrichtsorganisation und schulische Sozialisation. Weinheim: Beltz, 1971

HUSÉN, T.: Soziale Umwelt und Schulerfolg. Frankfurt/M: Diesterweg, 1977

INFRA-Test Sozialforschung: Gesamtschule im Flächenversuch. Ergebnisse einer Umfrage im Kreis Wetzlar. In: Gesamtschule, 1975 (Mai), 17-18

JOERESKOG, K.G.: A General Approach to Confirmatory Maximum Likelihood Factor Analysis. In: Psychometrika, 1969, 34, 183-202

JOERESKOG, K.G.: A General Method for Analysis of Covariance Structures. In: Biometrika, 1970, 57, 239-251

JOERESKOG, K.G. & van THILLO, M.: LISREL: A General Computer Programm for Estimating a Linear Structural Equation System Involving Multiple Indicators of Unmeasured Variables. In: Research Bulletin 72-56. Princeton, N.J.: Educational Testing Service, 1972

JOERESKOG, K.G. & SOERBOM, D.: LISREL III: Estimation of Linear Structural Equation Systems by Maximum Likelihood Methods. A Fortran IV Program. Chicago: National Educational Resources, 1976

KEIM, W. (Hg.): Gesamtschule. Bilanz ihrer Praxis. Hamburg: Hoffmann & Campe, 1973

KIM, J.-O.: Multivariate Analysis of Ordinal Variables. In: Amer. J. Sociol., 1975, 81, 261-298

KLAFKI, W. & STOECKER, H.: Innere Differenzierung des Unterrichts. In: Zt. f. Pädagogik, 1976, 22, 497-523

KLEINING, G. & MOORE, H.: Soziale Selbsteinstufung (SSE). Ein Instrument zur Messung sozialer Schichten. In: Köln. Zt. f. Soziol. Sozialpsych., 1968, 20, 502-522

KRAMP, W.: Studien zur Theorie der Schule. München: Kösel, 1973

LIENERT, G.A.: Testaufbau und Testanalyse. Weinheim: Beltz, 1961

LORD, F.M. & NOVICK, M.R.: Statistical Theories of Mental Test Scores. Reading (Mass.): Addison-Wesley, 1968

LOSER, F.: Methodische Differenzierung des Unterrichts durch Differenzierung von Unterrichtsmethoden. In: Bildung u. Erziehung, 1970, 23, 351-373

LUKESCH, H. & TISCHLER, A.: Kreuzvalidierung der Marburger Skalen zur Erfassung des elterlichen Erziehungsstils. In: Psychol. u. Praxis, 1975, 19, 19-30

MASENDORF, F. & TSCHERNER, K.: Aspekte des Interaktions- und Unterrichtsstils von Lehrern aus der Sicht der Schüler. In: Zt. f. Entwicklungspsychol. u. Pädag. Psychol., 1973, 5, 73-90

McDONALD, R.P.: A Unified Treatment of the Weighting Problem. In: Psychometrika, 1968, 33, 351-381

MERZ, F. & STELZL, I.: Einführung in die Erbpsychologie. Stuttgart: Kohlhammer, 1977

MUELLER, P.: Sind Jungen dümmer? In: Acta Paedapsychiatrica, 1970, 37, 326-336

NIE, N.H., HULL, C.H., JENKINS, J.G., STEINBRENNER, K. & BENT, D.H.: SPSS: Statistical Package for the Social Sciences. New York: McGraw-Hill, 1975

NIEHL, F.: Chancengleichheit ohne Chance? Stuttgart: Klett, 1975

NUNNALLY, J.C.: Psychometric Theory. New York: McGraw-Hill, 1967

OSTERLOH, H. & ROLAND, D.: Zu einigen Problemen des Lehrer-Schüler-Verhältnisses. In: Pädagogik, 1970, 25, (1. Beiheft), 18-32

OTTO, K.: Disziplin bei Jungen und Mädchen. Berlin: Deutscher Verlag der Wissenschaften, 1970

PEISERT, H. & DAHRENDORF, R.: Der vorzeitige Abgang vom Gymnasium. Villingen: Neckar-Verlag, 1967

POPPER, K.R.: Logik der Forschung. Tübingen: Mohr, 1969³

PRELL, S., SCHIEFELE, H. & ULICH, D.: Leistungsdifferenzierung und individuelle Förderung. München: Oldenbourg, 1972

PREUSS, O.: Soziale Herkunft und die Ungleichheit der Bildungschancen. Weinheim: Beltz, 1970

RANG, A. & SCHULZ, W. (Hg.): Die differenzierte Gesamtschule. München: Piper, 1969

RAO, C.R.: Estimation and Tests of Significance in Factor Analysis. In: Psychometrika, 1955, 20, 93-112

REVENSTORF, D.: Lehrbuch der Faktorenanalyse. Stuttgart: Kohlhammer, 1976

ROLLETT, B. & BARTRAM, M.: Konstruktion eines Prüfverfahrens der Anstrengungsvermeidung als Prognoseinstrument für den Lernerfolg. In: ECKENSBERGER, L.H. & ECKENSBERGER, U.S. (Hg.): Bericht über den 28. Kongress der Deutschen Gesellschaft für Psychologie 1972. Göttingen: Hogrefe, 1973, 95-105

ROLLETT, B. & BARTRAM, M.: AVT: Anstrengungsvermeidungstest. Braunschweig: Westermann, 1977 (1. Fassung 1972)

ROTH, W.: Pro und Kontra Gesamtschule. Strukturen für die Befürwortung und Ablehnung: Lehrer. In: Westermanns Pädag. Beitr., 1975, 75, 142-148

SCHEUCH, E.K.: Sozialprestige und soziale Schichtung. In: Köln. Zt. f. Soziol. Sozialpsychol., 1961 Sonderheft 5, 65-103

SCHNEEWEISS, H.: Oekonometrie. Würzburg: Physica, 1971

SCHOENBERG, R.: Strategies for Meaningful Comparison. In: COSTNER, H.L. (ed.): Sociological Methodology 1972. San Francisco: Jossey-Bass, 1972, 1-35

SCHORB, A.O. & SCHMIDBAUER, M.: Aufstiegsschulen im sozialen Wettbewerb. Stuttgart: Klett, 1973

SCHWARZER, C.: Kinderpersönlichkeit und Lehrerurteil bei Schulbeginn - Kognitive Stile, Fähigkeiten und sozio-ökonomischer Status als Einflussgrössen für schulischen Erfolg. Dissertation. Kiel: Philosophische Fakultät der Christian-Albrechts-Universität, 1975

SCHWARZER, C.: Schulversagen und Lehrerurteilsverhalten. In: Westermanns Pädag. Beitr., 1976, 28, 309-313 (a)

SCHWARZER, C.: Lehrerurteil und Schülerpersönlichkeit. München: Kösel, 1976 (b)

SCHWARZER, R. & ROYL, W.: Angst und Schulunlust als Sozialisationseffekte verschiedener Schularten. In: Zt. f. Pädagogik, 1976, 22, 547-557

SEWELL, H. & SHAH, V.P.: Parents' Education and Children's Educational Aspirations and Achievements. In: Amer. Sociol. Rev., 1968, 33, 191-209

SIMON, H.A. & ANDO, A.: Aggregation of Variables in Dynamic Systems. In: Econometrica, 1961, 29, 111-138

SOERBOM, D. & JOERESKOG, K.G.: COFAMM: Confirmatory Factor Analysis with Model Modification. A Fortran IV Program. Chicago: National Educational Resources, 1976

STAPF, K.H., HERRMANN, T., STAPF, A., & STAECKER, K.H.: Psychologie des elterlichen Erziehungsstils. Stuttgart: Klett, 1972

STARK, G.: Bringen Leistungskurse Chancengleichheit? In: Informationen zum Projekt Leistungsmessung in Gesamtschulen, Heft 3. Frankfurt/M.: Deutsches Institut für Internationale Pädagogische Forschung, 1971

STEINKAMP, G.: Lehrer voller Vorurteile. In: HIELSCHER, H. (Hg.): Die Schule als Ort sozialer Selektion. Heidelberg: Quelle & Meyer, 1972, 60-80

SUPPES, P. & ZINNES, J.L.: Basic Measurement Theory. In: LUCE, R.D., BUSH, R.R. & GALANTER, E. (eds.): Handbook of Mathematical Psychology, Vol. I. New York: Wiley, 1963, 1-76

TAUSCH, R. & TAUSCH, A.-M.: Erziehungspsychologie. Göttingen: Hogrefe, 1968[3]

TREIBER, B., WEINERT, F.E. & GROEBEN, N.: Bedingungen individuellen Unterrichtserfolgs. In: Zt. f. Pädagogik, 1976, 22, 153-179

TUKEY, J.W.: Causation, Regression, and Path Analysis. In: KEMPTHORNE, O., BANCROFT, T.A., GOWEN, J.W. & LUSH, J.L. (eds.): Statistics and Mathematics in Biology. Ames (Iowa): Iowa State College Press, 1954, 35-66

VOLLMER, F. & ALMAS, R.: Sex Differences in Achievement Motivation. In: Scand. J. of Psychol., 1974, 15, 310-313

WASNA, M.: Probleme der Gesamtschule - Aufgaben praxisbezogener psychologischer Forschung. In: NICKEL, H. & LANGHORST, E. (Hg.): Brennpunkte der pädagogischen Psychologie. Bern: Huber, 1973, 274-283

WEINER, B. & KUKLA, A.: An Attributional Analysis of Achievement Motivation. In: J. of Personal. and Social Psychol., 1970, 15, 1-20

WEISS, R.H.: Grundintelligenztest - Skala 2 (CFT2). Braunschweig: Westermann, 1972

WERRES, W.: Brennpunkte der Leistungsdifferenzierung des Unterrichts. In: WIEDERHOLD, K.A. (Hg.): Differenzierung in Schule und Unterricht. Ratingen: Henn, 1975, 43-59

WERTS, C.E., ROCK, D.A., LINN, R.L. & JOERESKOG, K.G.: Comparison of Correlations, Variances, Covariances, and Regression Weights with or without Measurement Error. In: Psychol. Bull., 1976, 83, 1007-1013

WILEY, D.E.: The Identification Problem for Structural Equation Models with Unmeasured Variables. In: GOLDBERGER, A.S. & DUNCAN, O.D. (eds.): Structural Equation Models in the Social Sciences. New York: Seminar Press, 1973, 69-83

WILHELM, T.: Theorie der Schule. Stuttgart: Metzler, 1967

ZEIHER, H.: Gymnasiallehrer und Reformen. Stuttgart: Klett, 1973

A N H A N G

Anhang 1: Instruktionen und Items

1. Allgemeine Einführung 152
2. Personalstatistische Fragen 155
3. Mutter-Unterstützungsskala 156
4. Vater-Unterstützungsskala 159
5. AVT zur Erfassung der Anstrengungsbereitschaft 162
6. Lehrer-Unterstützungsskala 165
7. Lehrer-Sympathieskala 168

Erläuterungen zu den Instruktionen und Items:

- Die allgemeine Einführung und die Instruktionen zu den Skalen wurden mündlich vorgetragen. Für den Schüler waren nur die Items und die personalstatistischen Fragen schriftlich (in "Heften") fixiert.
- Zu den personalstatistischen Fragen: In der ausgewählten Schule pflegte man die Kurseinstufungen durch römische Zahlen zu kennzeichnen (wobei: Kurs I = Kurs A, Kurs II = Kurs B, Kurs III = Kurs C).
- Bei jedem Item sind in Klammern die in dieser Untersuchung den Antwortmöglichkeiten zugeordneten Zahlenwerte aufgeführt (die in der Darbietung weggelassen waren). Welche Zahlenwerte die befragten Schüler im einzelnen erhielten, geht aus den Tabellen des Anhangs 2 hervor.

1. Allgemeine Einführung

(Begrüssung) Mein Name ist ...

Wir kommen von der Hochschule in Zürich (bzw. Giessen) und möchten hier eine Untersuchung durchführen. Wir möchten wissenschaftlich untersuchen, wie Schüler in eurem Alter über verschiedene Dinge denken. Mit unserer Untersuchung wollen wir dazu beitragen, dass ihr euch besser verstehen könnt, zum Beispiel: dass ihr versteht, warum ihr so geworden seid, wie ihr heute seid. Wir machen diese Untersuchung auch, weil wir erreichen wollen, dass ihr besser lernen und behalten könnt, und auch: dass ihr besser mit Eltern und Lehrern auskommen könnt.

Wir haben euch einige Hefte mitgebracht, in denen Fragen und Sätze stehen. Wir haben da zum Beispiel Sätze, die beschreiben, wie eure Lehrer zu euch sein könnten. Ihr sollt uns sagen, wie diese Sätze auf eure Lehrer zutreffen. Um euch diese Aufgabe zu erleichtern, haben wir uns zu jeder Frage und zu jedem Satz schon einige Antworten überlegt und für euch aufgeschrieben. Alles, was ihr nun tun müsst, ist, aus den von uns vorgeschlagenen Antworten diejenige Antwort herauszusuchen, die für euch am besten zutrifft. So einfach ist das.

Ihr braucht auch keine Angst zu haben, dass die Lehrer nachsehen, wie ihr geantwortet habt. Wir zeigen die Hefte niemandem in der Schule. Niemand ausser mir und meinen Mitarbeitern darf eure Hefte sehen. Nicht einmal eure Eltern dürfen eure einzelnen Antworten zu Gesicht bekommen. Ich sage euch dies allerdings nicht darum so deutlich, damit ihr jetzt meint, es würden nun ganz geheime Fragen gestellt - nein, überhaupt nicht deswegen. Ich möchte euch nur überzeugen, dass ihr ruhig so antworten könnt, wie euch der Schnabel gewachsen ist, dass ihr also ehrlich sagen könnt, was ihr denkt. Dass es uns da-

mit sehr ernst ist, euch davon zu überzeugen, könnt ihr ja schon daran sehen, dass wir es erreicht haben, dass kein Lehrer während der Untersuchung anwesend sein darf. Ausserdem haben wir erreicht, dass alle eure Hefte gleich nach der Untersuchung von uns wieder mitgenommen werden dürfen, ohne dass irgendeiner von der Schule nachsehen kann.

Darum sollt ihr auch nirgends euren Namen hinschreiben. Eure Namen sind für die Durchführung unserer Untersuchung völlig unwichtig. Nur: Um sicherzugehen, dass die verschiedenen Hefte in den Pausen und am Schluss nicht durcheinanderkommen, sollt ihr eure Schülernummern angeben. Wir wollen ja bei der Auszählung eurer Antworten die Möglichkeit haben zu sehen, welche Hefte zusammengehören. Schreibt daher bitte eure Schülernummer oben auf jedes neue Heft, das ihr bekommt. So braucht ihr euren Namen nicht zu nennen.

Die meisten Fragen werdet ihr sicher ohne Schwierigkeiten beantworten können. Wichtig ist dabei nur, dass ihr wirklich jede Frage der Reihe nach, also von der ersten bis zur letzten, beantwortet. Ueberlegt nicht erst, welche Antwort vielleicht den besten Eindruck machen könnte. Antwortet bitte so, wie es für euch persönlich zutrifft. Es braucht auch keiner auf das Heft des Nachbarn zu sehen; denn jeder denkt über die Fragen ja anders, und richtige oder falsche Antworten gibt es hier nicht. Uns kommt es vielmehr auf eure eigene Meinung an.

Es kann sein, dass ihr euch bei manchen Fragen gar nicht so recht entscheiden könnt zwischen zwei Antworten. In diesem Falle überlegt dann bitte noch einmal, ob nicht doch eine der beiden Antworten ein bisschen besser für euch zutrifft als die andere. Versucht möglichst alle Fragen zu beantworten. Es gibt auf jede Frage sicher immer eine Antwort, welche noch am ehesten für euch passt.

Noch einmal: Eure Teilnahme an der Untersuchung ist nur dann für die Wissenschaft und vor allem für euch selbst von Bedeutung, wenn ihr auch wirklich alle Fragen gewissenhaft und sorgfältig beantwortet.

Wenn ihr keine Fragen mehr habt, fangen wir sofort mit den ersten Fragen an.
(Vl teilt das Blatt mit den statistischen Fragen aus und sagt)
Ganz oben auf dem Blatt müsst ihr eure Schülernummer hinschreiben, die ihr von Herrn ... bekommen habt.

2. Personalstatistische Fragen

1. Schülernummer:

2. Alter in Jahren:

3. Junge Mädchen

   (Kreise bitte ein, ob Du ein Junge oder Mädchen bist)

4. Kurszugehörigkeit jetzt in:
   (Kreise den Kurs, in dem Du heute bist, bitte ein)

   (a) Mathematik: I
   II
   III

   (b) Englisch I
   II
   III

5. Kurszugehörigkeit zu Beginn des 5. Schuljahres:
   (Kreise den Kurs, in den Du zum ersten Mal vor zwei Jahren eingewiesen wurdest, bitte ein)

   (a) Mathematik I
   II
   III

   (b) Englisch I
   II
   III

3. Mutter-Unterstützungsskala

Das nächste Heft enthält Sätze, die beschreiben, wie eure Mutter zu euch sein könnte. Ihr sollt uns sagen, wie diese Sätze zutreffen. Zu jedem Satz haben wir uns schon einige Antworten überlegt. Um die Antwort, die am besten zutrifft, sollt ihr ganz deutlich einen Kreis machen.

Ich zeige euch an einem Beispiel, wie das gemacht wird.
(Vl hat den Satz mit den Antworten vorher schon an die Tafel geschrieben und liest Satz und Antworten laut vor)

"Meine Mutter geht sonntags mit mir spazieren"

sehr selten selten manchmal oft sehr oft

Sicher gibt es einige unter euch, die meinen, dass ihre Mutter sonntags manchmal mit ihnen spazieren geht. Da "manchmal" hier am besten zutrifft, machen diese einen Kreis um das "manchmal" (Vl demonstriert das an der Tafel). Diejenigen aber, die meinen, dass ihre Mutter sonntags sehr oft mit ihnen spazieren geht, machen um die Antwort "sehr oft" einen Kreis, weil diese Antwort für sie am besten zutrifft.

Jeder gibt also die Antwort, die für ihn am besten zutrifft. Wenn ihr dazu noch Fragen habt, dann meldet euch bitte.

So, und nun fangen wir an.

(Vl teilt Hefte aus und sagt) Es blättert aber noch keiner um, erst, wenn ich es sage. Schreibt bitte eure Schülernummern auf das leere Blatt (erste Blatt des Heftes) vor euch (Vl kontrolliert, geht durch die Reihen. Erst, wenn jeder seine Nummer genannt hat, fährt der Vl fort).

Ich werde die Sätze vorlesen. Es liest jeder still mit und überlegt, welche Antwort am besten für ihn zutrifft. Immer wenn ich den Satz vorgelesen habe, macht jeder einen Kreis

um die Antwort, die am besten zutrifft. Lasst bei keinem Satz die Antwort aus.

Wenn ihr keine Fragen mehr habt, dann blättern wir alle um und fangen an. (V1 beginnt den ersten Satz vorzulesen)

1. Meine Mutter sieht ein, dass ich recht habe, und gibt nach.

sehr oft	oft	manchmal	selten	sehr selten
(5)	(4)	(3)	(2)	(1)

2. Wenn ich etwas erzähle, hört meine Mutter aufmerksam und gespannt zu.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

3. Sie tröstet mich und hilft mir, wenn etwas schief gegangen ist.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

4. Wenn ich mit einem Freund Streit habe, sagt sie mir, wie ich mich verhalten soll.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

5. Sie bringt mir Dinge bei, die ich können möchte.

sehr selten	selten	manchmal	oft	sehr oft
(1)	(2)	(3)	(4)	(5)

6. Meine Mutter spricht mit mir über meine Freunde und sagt mir, wen sie nett findet.

sehr selten	selten	manchmal	oft	sehr oft
(1)	(2)	(3)	(4)	(5)

7. Wenn ich etwas getan habe, was sie falsch findet, darf ich sagen, warum ich es getan habe.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

8. Sie lobt mich, wenn ich mich wie ein Erwachsener benehme.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

9. Wenn ich etwas ausgefressen habe und ich mich dafür entschuldige, verzeiht sie mir schnell.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

10. Meine Mutter nimmt mich mit zu Gelegenheiten, zu denen andere Kinder nicht mitdürfen.

sehr selten	selten	manchmal	oft	sehr oft
(1)	(2)	(3)	(4)	(5)

11. Meine Mutter antwortet mir geduldig, wenn ich ihr viele Fragen stelle.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

12. Sie ermuntert mich, was ich noch nicht kann, selbst auszuprobieren.

sehr selten	selten	manchmal	oft	sehr oft
(1)	(2)	(3)	(4)	(5)

13. Wenn ich mich über etwas freue, merke ich, dass sie sich mit mir freut.

fast immer	oft	manchmal	selten	sehr selten
(5)	(4)	(3)	(2)	(1)

14. Wenn ich eine Aufgabe gut erledigt habe, belohnt mich meine Mutter.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

15. Sie lobt mich, auch wenn andere Leute dabei sind.

sehr oft	oft	manchmal	selten	sehr selten
(5)	(4)	(3)	(2)	(1)

4. Vater-Unterstützungsskala

Das nächste Heft enthält Sätze, die beschreiben, wie euer Vater zu euch sein könnte. Ihr sollt uns sagen, wie diese Sätze zutreffen. Um die Antwort, die am besten zutrifft, sollt ihr einen Kreis machen. Ihr wisst ja jetzt, wie das geht.

Wenn ihr dazu noch Fragen habt, dann meldet euch bitte.

So, und nun fangen wir an.

(Vl teilt Hefte aus und sagt) Es blättert aber noch keiner um, erst, wenn ich es sage. Schreibt bitte eure Schülernummern auf das leere Blatt vor euch.

(Vl kontrolliert, geht durch die Reihen. Erst, wenn jeder Schüler seine Nummer aufgeschrieben hat, fährt der Vl fort)

Ich werde die Sätze vorlesen. Es liest jeder still mit und überlegt, welche Antwort am besten für ihn zutrifft. Immer wenn ich den Satz vorgelesen habe, macht jeder einen Kreis um die Antwort, die am besten zutrifft. Lasst bei keinem Satz die Antwort aus.

Wenn ihr keine Fragen mehr habt, dann blättern wir alle um und fangen an. (Vl beginnt den ersten Satz vorzulesen)

1. Wenn ich eine gute Klassenarbeit geschrieben habe, bekomme ich von meinem Vater eine Belohnung.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

2. Er tröstet mich und hilft mir, wenn etwas schief gegangen ist.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

3. Wenn ich etwas gemalt oder gebastelt habe, kommt mein Vater und sieht es sich an.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

4. Mein Vater spricht mit mir über meine Freunde und sagt mir, wen er nett findet.

sehr selten	selten	manchmal	oft	sehr oft
(1)	(2)	(3)	(4)	(5)

5. Wenn ich etwas getan habe, was er falsch findet, darf ich sagen, warum ich es getan habe.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

6. Ich habe das Gefühl, dass ich mit ihm über alles sprechen kann, auch wenn ich etwas ausgefressen habe.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

7. Wenn ich etwas erzähle, hört mein Vater aufmerksam und gespannt zu.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

8. Er ermuntert mich, das, was ich noch nicht kann, selbst auszuprobieren.

sehr selten	selten	manchmal	oft	sehr oft
(1)	(2)	(3)	(4)	(5)

9. Wenn er mich bestraft, merke ich, dass er es nicht gern tut.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

10. Manchmal beobachte oder höre ich Dinge, die ich nicht verstehe; aber mein Vater erklärt sie mir dann.

fast immer	oft	manchmal	selten	sehr selten
(5)	(4)	(3)	(2)	(1)

11. Mein Vater sieht ein, dass ich recht habe, und gibt nach.

sehr oft	oft	manchmal	selten	sehr selten
(5)	(4)	(3)	(2)	(1)

12. Wenn mein Vater sich über mich freut, ist er sehr liebevoll und zärtlich.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

13. Wenn ich mich über etwas freue, merke ich, dass er sich mit mir freut.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

14. Wenn ich eine Aufgabe gut erledigt habe, belohnt mich mein Vater.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

15. Er lobt mich, auch wenn andere Leute dabei sind.

sehr oft	oft	manchmal	selten	sehr selten
(5)	(4)	(3)	(2)	(1)

5. AVT zur Erfassung der Anstrengungsbereitschaft

Das nächste Heft enthält Sätze, die euer Fühlen und Handeln betreffen. Ihr sollt uns sagen, ob diese Sätze allgemein für euch zutreffen oder nicht. Bei jedem Satz haben wir immer zwei Antworten vorgegeben: man kann antworten "stimmt" oder man kann antworten "stimmt nicht".

Ihr sollt die Sätze so beantworten, wie es für euch nicht nur heute, sondern auch sonst zutrifft. Wenn ihr meint, dass der Satz für euch zutrifft, dann macht ihr einen Kreis um das "stimmt" (Vl demonstriert dies an der Tafel. Dort stehen nur "stimmt" und "stimmt nicht"). Wenn der Satz für euch nicht passt, dann müsst ihr einen Kreis um das "stimmt nicht" machen (Vl zeigt dies an der Tafel).

Wenn ihr dazu noch Fragen habt, dann meldet euch bitte.

So, und nun fangen wir an.

(Vl teilt Hefte aus und sagt) Es blättert aber noch keiner um, erst, wenn ich es sage. Schreibt bitte eure Schülernummern auf das leere Blatt vor euch (Vl kontrolliert, geht durch die Reihen. Erst, wenn jeder Schüler seine Nummer aufgeschrieben hat, fährt der Vl fort).

Ich werde die Sätze vorlesen. Es liest jeder still mit und überlegt, welche Antwort am besten für ihn zutrifft. Immer wenn ich den Satz vorgelesen habe, macht jeder einen Kreis um die Antwort, die am besten zutrifft. Lasst bei keinem Satz die Antwort aus.

Wenn ihr keine Fragen mehr habt, dann blättern wir alle um und fangen an. (Vl beginnt den ersten Satz vorzulesen)

1. Meine Mutter will immer gerade dann, dass ich ihr helfe, wenn eine gute Sendung im Fernsehen läuft.
   stimmt (0) stimmt nicht (1)

2. Es passiert mir oft, dass ich gar nicht höre, wenn der Lehrer uns Aufgaben gegeben hat.
   stimmt (0) stimmt nicht (1)

3. Ich habe keine Lust mehr zu arbeiten, wenn ich an das Spielen denke.
   stimmt (0) stimmt nicht (1)

4. Wenn man dafür nicht so viel arbeiten müsste, wäre ich gerne ein guter Schüler.
   stimmt (0) stimmt nicht (1)

5. In der Schule werde ich oft nicht fertig, weil mein Nachbar mich stört.
   stimmt (0) stimmt nicht (1)

6. Wenn ich arbeite, muss ich immer daran denken, was ich jetzt spielen könnte.
   stimmt (0) stimmt nicht (1)

7. Mit meinen Aufgaben werde ich oft nicht fertig, weil ich dazwischen zu lange nachdenken muss.
   stimmt (0) stimmt nicht (1)

8. Ich kann nichts dafür, wenn ich das, was ich auswendig lernen soll, nicht behalten kann.
   stimmt (0) stimmt nicht (1)

9. Mein Lehrer (meine Lehrerin) sagt oft, dass ich schneller arbeiten soll.
   stimmt (0) stimmt nicht (1)

10. Wenn draussen die Sonne scheint, kann ich einfach nicht arbeiten.
    stimmt (0) stimmt nicht (1)

11. Ich habe oft Langeweile.
    stimmt (0) stimmt nicht (1)

12. Warum soll ich am Wochenende lernen, wenn die anderen Leute auch nicht arbeiten.
    stimmt (0) stimmt nicht (1)

13. Ich verstehe überhaupt nicht, warum man das Einmaleins auswendig lernen soll.
    stimmt (0) stimmt nicht (1)

14. Die Lehrer verwenden zu viele Wörter, die ich nicht verstehe.
    stimmt (0) stimmt nicht (1)

15. Wenn ich keine Hausaufgaben aufbekäme, könnte ich in der Schule viel besser aufpassen.
    stimmt (0) stimmt nicht (1)

16. Ich kann meine Aufgaben nicht ordentlich machen, weil ich immer gestört werde.
    stimmt (0) stimmt nicht (1)

17. Rechtschreibung ist nicht wichtig, man könnte ja auch alles ganz anders schreiben.
    stimmt (0) stimmt nicht (1)

18. Was mich nicht interessiert, will ich gar nicht machen.
    stimmt (0) stimmt nicht (1)

19. Wenn ich lange schreiben muss, werde ich ganz müde.
    stimmt (0) stimmt nicht (1)

20. Der Lehrer (die Lehrerin) nimmt mich nie dran, wenn ich aufzeige, warum soll ich mich dann melden.
    stimmt (0) stimmt nicht (1)

6. Lehrer-Unterstützungsskala

Das nächste Heft enthält Sätze, die beschreiben, wie eure Lehrer zu euch sein könnten. Ihr sollt uns sagen, wie diese Sätze auf eure Lehrer zutreffen. Um die Antwort, die am besten zutrifft, sollt ihr einen Kreis machen. Von den vorigen Heften her wisst ihr nun schon, wie das gemacht wird. Deshalb brauche ich jetzt auch kein Beispiel zu bringen.

Wichtig ist, dass jeder die Antwort gibt, die für ihn am besten zutrifft. Ueberlegt nicht erst, welche Antwort vielleicht den besten Eindruck machen könnte. Jeder erlebt die Lehrer anders, und ihr wählt einfach die Antwort aus, die eurer eigenen, persönlichen Meinung am nächsten kommt. Ihr braucht ja keine Angst zu haben, dass die Lehrer nachsehen, wie ihr geantwortet habt.

Wenn ihr dazu noch Fragen habt, dann meldet euch bitte.

So, und nun fangen wir an.

(Vl teilt Hefte aus und sagt) Es blättert aber noch keiner um, erst, wenn ich es sage. Schreibt bitte eure Schülernummern auf das leere Blatt vor euch (Vl kontrolliert, geht durch die Reihen. Erst, wenn jeder Schüler seine Nummer aufgeschrieben hat, fährt der Vl fort).

Ich werde wieder die Sätze vorlesen. Es liest jeder still mit und überlegt, welche Antwort am besten für ihn zutrifft. Immer wenn ich den Satz vorgelesen habe, macht jeder einen Kreis um die Antwort, die am besten zutrifft. Lasst bei keinem Satz die Antwort aus.

Wenn ihr keine Fragen mehr habt, dann blättern wir alle um und fangen an. (Vl beginnt den ersten Satz vorzulesen)

1. Wenn ich traurig bin, fragen mich meine Lehrer nach dem Grund.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

2. Wenn ich etwas nicht kann, machen sie mir Mut, es zu versuchen.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

3. Wenn ich eine Aufgabe gut erledigt habe, loben mich meine Lehrer.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

4. Wenn meine Lehrer mich bestrafen, erklären sie mir, warum sie es tun.

sehr selten	selten	manchmal	oft	sehr oft
(1)	(2)	(3)	(4)	(5)

5. Wenn es mir nicht gut geht, nehmen meine Lehrer Rücksicht darauf.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

6. Meine Lehrer freuen sich, wenn ich anderen in der Klasse helfe.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

7. Meine Lehrer sorgen dafür, dass wir uns in der Schule wohl fühlen.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

8. Wenn ich einmal eine schlechte Klassenarbeit geschrieben habe, kann ich meinen Lehrern sagen, wie das gekommen ist.

sehr selten	selten	manchmal	oft	sehr oft
(1)	(2)	(3)	(4)	(5)

9. Es kommt vor, dass sie mich vor der ganzen Klasse loben.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

10. Wenn ich Sorgen habe, kann ich meine Lehrer um Rat fragen.

fast immer	oft	manchmal	selten	sehr selten
(5)	(4)	(3)	(2)	(1)

11. Wenn mich meine Lehrer bestrafen, merke ich, dass sie es nicht gern tun.

sehr oft	oft	manchmal	selten	sehr selten
(5)	(4)	(3)	(2)	(1)

12. Wenn ich einmal eine schlechte Note bekommen habe, beruhigen mich meine Lehrer und machen mir Mut.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

13. Wenn ich eine gute Klassenarbeit geschrieben habe, loben mich meine Lehrer.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

14. Wenn ich mich mit Klassenkameraden einmal zanke, versuchen meine Lehrer den Streit gerecht zu schlichten.

fast immer	oft	manchmal	selten	fast nie
(5)	(4)	(3)	(2)	(1)

15. Wenn ich bei einer Aufgabe nicht weiter eiss, helfen mir meine Lehrer.

sehr oft	oft	manchmal	selten	sehr selten
(5)	(4)	(3)	(2)	(1)

7. Lehrer-Sympathieskala

(Die folgenden Items wurden in der Darbietung direkt an die Items der Lehrer-Unterstützungsskala angeschlossen; daher entfällt hier die Instruktion)

1. Meine Lehrer schenken mir volles Vertrauen.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

2. Meine Lehrer sind richtig gemein zu mir.

fast nie	selten	manchmal	oft	fast immer
(5)	(4)	(3)	(2)	(1)

3. Meine Lehrer würden es nicht gerne sehen, wenn ich Klassensprecher würde.

fast immer	oft	manchmal	selten	fast nie
(1)	(2)	(3)	(4)	(5)

4. Ich fühle mich von meinen Lehrern gerecht behandelt.

fast nie	selten	manchmal	oft	fast immer
(1)	(2)	(3)	(4)	(5)

5. Meine Lehrer finden mich unsympathisch.

fast immer	oft	manchmal	selten	fast nie
(1)	(2)	(3)	(4)	(5)

Anhang 2: Information zu Daten:

Die Daten der Untersuchung können vom Autor angefordert werden (Adresse: Psychologisches Institut der Universität Zürich/Schweiz).

Lebenslauf

1942	Geboren am 6. 1. 1942 in Berlin
1948 - 1951	Besuch der Volksschule in Bergen bei Celle
1951 - 1953	Besuch der Volksschule in Hattingen/Ruhr
1953 - 1963	Besuch des Evangelischen Stifts-Gymnasiums in Gütersloh. Dort bestand ich am 7. 2. 1963 das Abitur.
1963 - 1964	WS 63/64 und SS 64 Studium der Psychologie an der Universität in Hamburg.
1964 - 1968	WS 64/65 bis SS 68 Studium der Psychologie an der Universität in Marburg/Lahn. Dort bestand ich am 30. 12. 1968 die Diplom-Hauptprüfung für Psychologen.
1969 - 1972	Forschungsassistent in der Sektion "Sozialpsychologie" der Abteilung Psychologie am Max-Planck-Institut für Psychiatrie in München.
1972 - 1974	Assistent des Schweizerischen Nationalfonds, an der Sozialforschungsstelle des Psychologischen Instituts der Universität Zürich
Seit 1974	Seit 1. 3. 1974 wissenschaftlicher Assistent beim Psychologischen Institut der Universität Zürich.